

nature

26 January 1978

Science in the EEC still a problem

If there is one thing on which all involved European scientists and administrators agree, it is that a research and development policy for the European Economic Community is proving a mighty difficult thing to bring forth. There are many reasons for this. One is that there have long been doubts that the community of nine nations should be concerned other than peripherally in basic research—institutions such as the European Science Foundation, the European Molecular Biology Laboratory and CERN draw from a wider range of nations. Another is that applied science and technology policy is difficult to devise in the absence of a broad community-wide economic policy and the sort of political union which makes British taxpayers less agitated if the R & D funds they provide lead to the establishment of new industry in, say, Italy. Finally, the policy must take note of already existing community laboratories at Geel, Petten, Ispra and Karlsruhe—the so-called Joint Research Centre—a hangover from Euratom days when there was a naive and erroneous belief that the then community members, already lavishly equipped with their own nuclear research laboratories, would have work for yet another lavishly-equipped and wide open facility.

Add to this differences in the aspirations of larger and smaller nations and the tiny dimensions of the community's R & D budget in comparison with those of individual nations (1½% of the combined non-defence R & D budget of member states), and you might reasonably wonder whether it was worth the community persisting in trying to keep a scientific programme going. In July 1977, however, the Commission of the Community had a go at bringing some new order into the science and technology programme, with a document devoted to intentions for 1977-80 and draft resolutions and decisions embodying these intentions (*Nature* 268, 96). The House of Lords Select Committee on the European Communities has been considering the evolving policy and hearing evidence from scientists with European interests. Their report is now published (Commons Paper 37; £2.60). For those who imagine that any committee of Lords could hardly get to grips with such an esoteric matter, it should be added that the committee comprised several scientists and engineers, including Lords Ashby, Hinton and Zuckerman.

It was perhaps unfortunate that the committee had completed its gathering of evidence before the Commission's report was published, for although there were some interesting presentations on the European dimension to science and a particularly spirited contribution from Professor Pierre Aigrain, it is difficult to see some of the committee's conclusions on the community document emerging very clearly from their earlier deliberations; indeed one draft decision on industrial research which the committee commends does not seem to have been the subject of any discussion whatsoever. Even so, it is possible to discern from the general drift of the evidence that the community is of most value when it is used simply to co-ordinate national research ('concerted action'), some value in financial support of domestic projects of community-wide interest ('in-

direct action') and at its most problematical when the research is done in the Joint Research Centre ('direct action'). Not least of the problems in the last case is the bureaucratic constraint imposed on the actions of the director at Ispra which gives him less freedom than a laboratory director might reasonably expect. The freedom of Sir John Kendrew at EMBL to run his own show was not lost on the committee.

What, then, to do about the JRC? The community's document recommends, somewhat bravely, a mixed diet of nuclear safety, new energy sources, environment, resources and services. The Lords' committee suggest simply a "useful role in carrying out work not done elsewhere and, for example, 'ungrateful' research"—the latter being long-term unspectacular research. This, if anything, seems a recipe for even more dissatisfaction; a huge research laboratory devoted to unwanted projects is hardly likely to find a director able to give it even nominal cohesion or to prevent it from slipping rapidly into oblivion.

It is, however, the committee's views on forecasting which leave most to be desired. Some years ago the community established a group called Europe+30 to look into whether there should be long-term forecasting capacity, together with an Office of Technology Assessment, in the EEC. The group, under Lord Kennet, made its report in 1975 (now published in edited form as *The Futures of Europe* by C.U.P.). Its recommendations included the establishment of a permanent unit of between 30 and 70 people looking at all issues of concern to the community. Lord Kennet had a tough time when he appeared before the committee, which on that day notably lacked the scientists mentioned earlier. He or his team were accused of misuse of the English language, superficiality, amateurism, exceeding their brief and so on. But after the committee had heard from Lord Kennet and before the report was written, the community made its own move on forecasting. It proposed an attenuated version of Europe+30 called FAST—a forecasting-and-assessment-in-science-and-technology programme. This, as its name implies, would be restricted to one sector and would accordingly have a much smaller staff, probably of ten.

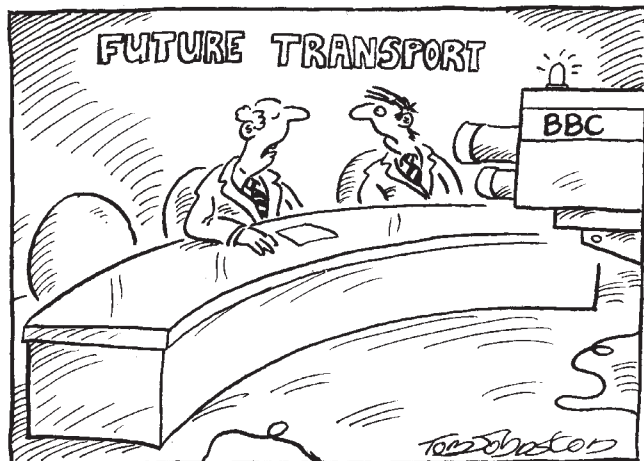
The Lords' committee, noticeably sceptical towards Europe+30, claims that FAST will be too diluted to be of value to the community. This is not argued out at all, and indeed could conceivably have gone the other way—that work in one sector would be more concentrated and hence of more value to the community. What is suggested in place of FAST is a scientific adviser, or team, reporting direct to the President. This is difficult to take seriously, particularly coming from Britain where, for better or worse, a scientific adviser reporting direct to the Prime Minister no longer exists.

Whatever happens in Brussels, it is time now for the Council of Ministers to take firm action. As the Lords remark, the ultimate responsibility for better organisation rests with ministers, and they had best wade into this problem with some vigour. □



Forecasting future transport

J. I. Gershuny, of the Science Policy Research Unit, University of Sussex, discusses the recommendations of the Department of Transport report of the Advisory Committee on Trunk Road Assessment (HMSO, January 1978)



"We were due to have a spokesman from TRRL here tonight ... but he's been held up in the traffic!"

OFFICIAL procedure in the UK for assessing trunk road construction schemes has, in recent years, been the subject of increasing criticism. Statisticians and transport experts have claimed that the procedures are technically faulty. Local objectors to road building schemes complain that the obscurity of the methods prevents their participation in the planning process. The Leitch Committee, which reported recently, supported the technical criticisms, but did not suggest much practical help for protesters.

Decisions to build trunk roads in the UK are based on two technical procedures. One is a forecasting exercise carried out by a group at the Transport and Road Research Laboratory (TRRL) in Buckinghamshire. The other is a computerised cost benefit analysis package called COBA, developed by the UK Department of Transport and its predecessors. Both have come under increasing criticism.

The TRRL forecasting procedure is described in the report LR650 published by the laboratory. This document makes forecasts of vehicle numbers extending into the quite distant future—the final time horizon is the year 2010. It puts forward three alternative estimates of the car population of the UK at this date—25.0, 25.9 and 26.2 million cars, or 0.43, 0.44 and 0.45 cars per head of population respectively (LR650, p15). These alternative estimates are very close together by the normal standards of social forecasting; can we really be this certain?

The reason for the small range of the alternative forecasts is to be found in the forecasting procedure adopted by the TRRL group. The predictions are based on the assumption that the pattern of growth is logistic (S-shaped) over time. A logistic growth model tends towards a given saturation level at a rate determined by some causal variable or variables—in the TRRL case, national income, motoring costs and time. So the high estimate of 26.2 million cars in 2010 involves an assumption of high national income growth and low growth in motoring costs, the low 25.0 estimate assumes low income and high costs growth, and the middle one involves medium costs and growth.

A first reason that these estimates should bunch so closely together is that the very steady historical pattern of growth of vehicle numbers in the fifteen years preceding the fore-

cast showed little fluctuation with variations in incomes and costs. So when calibrating the model, the future effect of variations of prices and incomes is similarly assumed to be small in relation to the effect of the passage of time. Even large assumed fluctuations make little difference to the forecast—because on past evidence the price and income elasticities of demand for private cars is low. Sceptics doubt that historically derived elasticities necessarily hold for the future.

A second reason for the bunching is that built into the structure of the model is a tendency for the effect of the variables on the forecast to be progressively suppressed over time as the value of the saturation asymptote is approached. The saturation level assumed in LR650 is 0.45 cars per head, and since the final predictions are so very close to this we might reasonably assume that the causal variables are contributing very little to the final result. Finally, the saturation level itself is very much a matter of speculation. In spite of sophisticated presentation, the results of the exercise come down to little more than guesses.

The cost-benefit assessment package, COBA, is rather more straight-forward, being, of course, subject to all the standard criticisms of cost-benefit analysis. The system requires that planners put money values on monuments and artifacts that have no natural economic worth, but nevertheless have considerable cultural value. Furthermore, cost-benefit analyses inevitably leave some costs completely unconsidered; COBA itself takes virtually no account of environmental damage caused by road building schemes. And even were all conceivable impacts taken into account, the effect of using a market-based system of valuation is that the preferences of the rich are given much more weight than those of the poor in the assessment.

Further COBA's assessment of future benefits is based in part on the predictions of future car numbers—giving rise to the suspicion that the TRRL vehicle forecasts may be self-fulfilling. However arbitrary the original predictions are, if, on the basis of these forecasts, new roads are built, and if, as many people suspect, the number of cars increases to fill the available road space, then the volume of traffic may well grow to the forecast level—whatever that level is! Hardly a rational way of formulating policy.

It was this sort of criticism that led to the appointment of the committee of enquiry into trunk road assessment under the chairmanship of Sir George Leitch. The committee's report contains an extremely thorough and lucid review of the assessment methods as currently practised. It examines vehicle forecasting procedures and methods of economic and environmental evaluation and compares them with practices in France, the USA and West Germany.

The Leitch report recommends that the current vehicle forecasting methods should be scrapped. To replace them the committee suggests a modelling procedure that gives much more weight to those causal factors—such as household organisation, geographical location of facilities and the availability of public transport—which actually determine people's desires to own and use cars. Such models would in particular be sensitive to the effects of public policies altering the balance of costs and convenience between public and private transport; some observers suggest that encouraging the use of public transport relative to private might in the long term have an important effect on the growth of car numbers, but the present forecasting procedure is quite insensitive to such changes in policy. This sort of conditional information about the effects on vehicle use of alternative feasible public policies would be much more useful to policymakers than the present unconditional forecasts. The report—as far as it goes—is to be welcomed.

□

Killer satellites

Farooq Hussain traces the history of the Soviet and US development of killer satellites

MILITARY operations depend heavily on satellites for intelligence gathering, navigation, weather information, command, control and communication. So their vulnerability to interference or destruction is a serious and legitimate concern of military planners. Satellites are delicate devices. To make one ineffective, it is necessary only to impair the performance of its sensitive equipment: its total destruction is unnecessary. As to means, radiation directed against a satellite can cause damage at much longer ranges than the effects of blast and flying debris from an explosion in the near-vacuum of space where shock waves have little effect. Nuclear warheads would be very effective against satellites—because of their lethal radiation. But the effects of nuclear anti-satellite weapons would be indiscriminate and would more than likely result in the destruction of friendly satellites as well as enemy ones. Killer satellites exploit the poor resistance of satellite components—particularly solar cells—to intense heating and radiation damage. High energy lasers can be easily directed at satellites, so their imagined use in killer satellites is extensive. Another possibility arises from the tendency satellites have to build up high electrostatic charges due to the higher electron density around the dark side of the planet. The use of ion beams directed at target satellites can cause arcing and discharge through the instrumentation either destroying or seriously damaging it. Precision guided missiles also offer the possibility of destroying satellites either by collision or by the use of a conventional warhead detonated close to the target.

The United States and the Soviet Union have signed separate treaties in 1963 and 1967 first prohibiting testing and then deployment of nuclear weapons in space. But since 1968 the Soviet Union has been conducting tests within the general Cosmos series of satellites and these have been widely interpreted by western observers as an anti-satellite development programme. In these tests an orbiting interceptor is manoeuvred so as to make one or more close passes by a target satellite. Out of the 27 such Cosmos launches that have taken place to date only seven have ended with the explosion of the interceptor and these explosions have not always been in the general vicinity of the target satellite. The latest test in which the interceptor exploded in orbit

occurred in December 1977. Other tests have taken place which have suggested that new manoeuvring techniques have been adopted for the interceptor satellites.

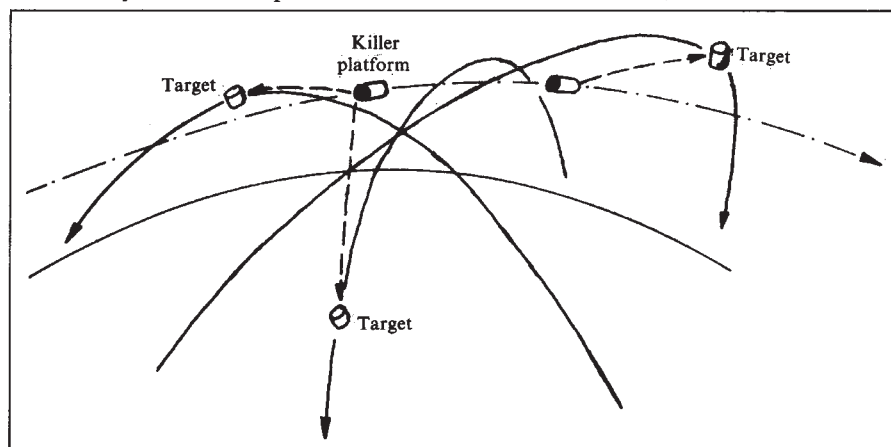
American concern over this satellite programme has arisen because of the sudden resumption of testing after a lapse of four years, and the accelerated rate at which testing has been taking place. One explanation for the Soviet tests is the development of reconnaissance satellites by China. Soviet testing of the Cosmos series appears to have followed Chinese reconnaissance satellite development since 1970. In addition the orbits followed by the Cosmos series seem more appropriate to interception of Chinese satellites than American ones.

American test

The only American test of an anti-satellite system took place in 1963

mechanical malfunction in deploying the payload. The programme was finally abandoned because of anticipated difficulties in getting the interceptor close enough to the target for conventional warheads to be effective.

Recent advances in the available energy, frequency range, and compactness of military lasers have prompted ideas that killer satellites may be able to carry high energy continuous wave or pulsed lasers to destroy target satellites. Ground-based lasers have been in use for a few years to interrogate reconnaissance satellites. The laser beam is used to illuminate the satellite, and the spectral pattern of the reflected laser radiation can be used to interpret various characteristics of the target. Using this technique it has been possible to determine the kind of equipment on the satellite and its sensitivity. Rumours that the Soviet use of lasers in this manner had tem-



The American system being developed by the Vought Corporation of Dallas, Texas

when a modified McDonnell Douglas Thor missile was fired against a missile booster in low earth orbit as the target. The Thor missile was brought within the range of the target for a simulated nuclear kill but the programme was dropped following the treaties banning nuclear weapons testing and weapons deployment in space. In a follow-on project the Vought Corporation built four infra-red homing guided missiles for use with conventional warheads. These missiles were to have been launched against satellites by a booster in a direct ascent trajectory towards the target without completing an orbit of the earth. Since the missiles were to have been armed with conventional warheads they needed a much smaller kill radius than for nuclear warheads and hence greater accuracy in guidance. But the first test launch suffered a booster failure and the second a

porarily blinded an American reconnaissance satellite last year were persistent in spite of official US Department of Defense denials. However general concern within the US Department of Defense and Congress has resulted in substantial increases in the American space defence budget. \$61 million was authorised for FY 1977 with \$126 million requested for FY 1978 and an estimated \$265 million being requested for FY 1979.

In September the United States Air Force announced the award of a \$58.7 million contract to the Vought Corporation of Dallas, Texas for the development of a killer satellite. The project developed out of a competition begun in 1975 between General Dynamics/Pomona and Vought to produce an unarmed miniature homing missile designed to destroy satellites by force of impact on collision. The

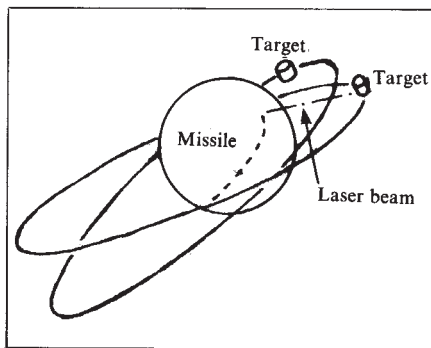
missile, about twenty centimetres in diameter and thirty centimetres long, is guided either by active radar or by far-infra-red seekers. Infra-red has special advantages because satellites give distinct thermal 'fingerprints'. More than ten homing missiles are carried into orbit by a highly manoeuvrable platform whose orbit is directed so as to provide the opportunity to fire missiles at the maximum number of target satellites within a minimum number of orbits. Free-fall tests of the infra-red guidance system have demonstrated mid-air activation, target acquisition and guidance to impact with a target. Electronic counter-measures against infra-red seekers (which are passive detectors, emitting no signal of their own) are more difficult than for active radar; and it is difficult for satellites to make rapid evasive manoeuvres in orbit. A similar system involving missiles armed with conventional warheads has also been under investigation.

Interpreting Soviet tests

By comparison, Soviet tests have only obliquely suggested a preference for destruction by large explosions in the vicinity of the target satellite. Since Soviet tests have never resulted in the physical destruction of a target, the explosion of the interceptor could be explained as a test of the vulnerability of the target to attack—rather than demonstrating the potential of the interceptor to destroy target satellites. The difficulty in interpreting properly the explosions in the Cosmos series also lends support to the idea that Soviet killer satellites may soon carry laser or ion-beam weapons. In any case the Soviet and American hunter-killer satellite developments should not be directly compared because each Super Power has very different approaches to the development of major weapon systems.

Presently US space defence research is concentrated upon problems of hardening satellites against attack and increasing the resistance of communication, command, and control transmissions to jamming. These battlefield transmissions now rely almost

completely on satellites. The third generation of American military communications satellites now being developed employ a technique of using multiple directional beams for transmission, and of shifting frequencies in the ultra and super high frequency bands (7 to 8 GHz) to avoid the effects of jamming. Satellite antennae are being developed that use electronic counter-measures to eliminate the effect of a directional jamming beam fixed on the satellite. Solar cells are being designed and built to withstand the effects of intense heating and radiation, by using filter coatings which do not absorb energy at typical laser wavelengths, and by using construction techniques with materials with high resistance to heat and radiation damage. The vulnerability of solar cells has led to an increased interest in nuclear power sources for satellites which until now have always



The Soviet system can only knock out one satellite at a time making it ineffective against the US global positioning array of 36 satellites

had the disadvantage of being much heavier and considerably more costly to operate. Three designs are currently under consideration all of which use P^{238} isotopes as the power source. With the considerable improvements in performance that are being achieved for satellites military dependence on them can be expected to increase throughout the 1970s and early 1980s. During the first half of the 1970s military missions accounted for about half of the American satellite launches and both the number of military satellites and their relative size and payload will increase further as the NASA Space Shuttle

begins operation.

The development of killer satellites is unlikely to provide any significant strategic advantage for either Super Power and the race to develop them is likely to be very costly. A surprise attack is hard to achieve with killer satellites because the preliminary manoeuvres are easily detected by other satellites. It is also possible to place in distant orbits satellites which are ready to replace ones destroyed by an attack. These so-called 'dark' satellites have characteristics such as very small radar cross-sections and low reflectivity which make them very difficult to detect and track. Because of their distant orbits they are relatively safe from attack in the first instance. The real advantage of destroying military satellites in a strategic nuclear conflict would appear only if one side destroyed a significant number of the opponent's satellites while retaining its own. A rapid destruction of all satellites would yield no advantage to either side.

It is perhaps to avoid a costly arms race that both the Soviet Union and the United States are anxious to reach an agreement limiting the development of killer satellites. The United States was reported to have requested Soviet consideration of such an agreement last March. The Soviet response was favourable and the United States National Security Council are presently considering a formal policy based on reports by various departmental agencies on the implications and verification problems of a hunter-killer satellite limitation or ban. The intention is to provide for an agreement separate to SALT and talks on a possible agreement over limiting killer satellite development may begin this Spring. Meanwhile US Defense Department officials are concerned that present programmes should not be effected by the agreement and it is unlikely that the incentive given to US space defence research by the recent Soviet testing of the Cosmos satellites will be abated by a limitation—unless the form of the agreement is radically different to those which have been made in other areas of strategic arms control. □

1979 looks like a good year for US science

CLIMATOLOGY, reproductive and developmental biology, and military-related basic research are three areas of science likely to receive a major boost from President Carter's Budget for the fiscal year 1979, which was submitted to Congress on Monday.

Other areas to benefit from a budget that proposes an overall increase of 10.9%—about 5% above expected in-

flation levels—for basic science include solar studies, which is given the green light for two major projects, and earthquake research. Potential casualties include one of the five planned space shuttles, and the Clinch River liquid metal fast breeder reactor demonstration project. The proposed termination of the Clinch River project would save about \$150 million, and would

result in virtually no increase in the energy research and development budget overall.

In general, the budget carries an optimistic message for basic science, in line with the results of a broad-ranging survey of the nation's research activities carried out at the President's request by the Office of Science and Technology Policy (OSTP). The projected

increase in spending on basic research is precisely the level recommended by the President's Science Adviser, Dr Frank Press, who claims "100% support" from the Office of Management and Budget.

But if previous budgets could be described as Keynesian, this one returns to a more classical, monetarist mould. Emphasis in funding is shifted away from development projects (for example in solar heating technology) which are felt to be the responsibility of the private sector and its market forces, back to more basic "long term" considerations.

Thus within an overall increase in support for research and development of 6.1%, to a total of \$27,900 million, the main increase will take place in basic science, from \$3,288 million in 1978 to \$3,647 million in 1979. In contrast the budget for applied research will increase by only 7.4% from \$6,248 million to \$6,711 million, and for development by an even lower rate of 4.6%, from \$16,753 to \$17,532 million.

A major beneficiary of the increased funding for basic research will be universities and colleges, many of whose problems—such as the obsolescence of equipment and the lack of opportunities for young scientists—were brought to the President's attention by the OSTP survey. Total R & D support for colleges is therefore suggested to rise from \$3,300 million in 1978 to \$3,561 million in 1979, an overall increase of 7.9%, with particularly large increases in funds for agricultural (10.7%), energy (11.7%), and military research. (In the light of the benefits which these increases are expected to bring, the NSF's proposed basic research stability grants, whose \$4.5 million was "impounded" by the President from the 1978 budget a few months ago, will be permanently dropped.)

One area to receive a major boost is climate research. A broad initiative in this area, involving eight agencies and a 39% increase in funding to \$104 million, will be coordinated by the National Oceanic and Atmospheric Administration of the Department of Commerce.

Three important aspects of the proposed 1979 budget will be: the development of satellite sensors for ozone and earth radiation monitoring—a major field study of the equatorial Pacific—and basic research on carbon dioxide levels in the atmosphere and oceans.

Two new programmes are included in the budget of the National Aeronautics and Space Administration (NASA). The first of these is the Solar Polar Mission, a joint project with the European Space Agency to send two space craft, launched from the Space Shuttle in 1983, past Jupiter, in an orbit that will permit the first study of

the polar regions of the sun. And the second project is the Solar Mesospheric Explorer, which will permit study of the effects of solar radiation on the earth's ozone layer.

At the Department of Health, Education and Welfare, whose basic research budget is proposed to be increased by 14.9% to \$992 million, areas scheduled for increased support include reproduction and family planning, developmental biology, and child development, as well as the potential environmental hazards to human health.

There are bright prospects, too, for military-related basic research. Spending by the Department of Defense is proposed to increase by 14%, almost as much as for health research, reflecting the view of Defense Secretary Dr Harold Brown that this area has been underfunded in recent years.

One result of this will, according to Dr Press, be the re-establishment of connections between the military establishment and the university research community which, he says, "have been allowed to lapse in recent years" and the R & D budget of the Department of Energy includes provision for work on the controversial neutron bomb.

The future for basic research—at least in most areas—therefore looks better than it has for several years past. But the *leit-motif* remains that of the ultimate pay-off. Dr Press, pointing out that the applications of basic science may take 10 or 20 years to come through, characterises this as an area in which industry tends as a consequence to "underinvest", thus justifying increased federal expenditure. And Dr Robert Frosch, administrator of NASA, ended his statement on the Agency's budget with the words that it would "give the American taxpayer the return on his investment he expects and deserves".

This approach is also reflected in the budget of the NSF. One area due for significant expansion, for example, is the integrated basic research programme of the new Applied Science and Research Application Directorate.

This is scheduled for an increase from \$1.9 to \$7.0 million, and will be concerned with "identifying basic research relevant to significant national problems", jointly with the basic research directorate of NSF.

One area to have suffered from this cost-conscious approach is high energy physics, whose budget is planned to expand by only 6.7%. The budget allows substantial sums for the construction of the intersecting storage accelerator (Isabelle) at Brookhaven, and for a new high-intensity beam facility at the Lawrence Berkeley Laboratory.

At Fermilab in Illinois, however, although the energy saver/doubler project is being formally moved from the R & D to the construction phase, only \$10 million of the proposed authorisation of \$38.9 million for construction is scheduled to be spent in 1979—and this will be partially met by reducing the \$86.1 million authorised for running costs in 1978 and \$81.8 million authorised in 1979.

But if the cost/benefit consciousness that runs through the proposed budget has resulted in decisions that are unwelcome in some quarters, it is likely to increase the chances of the budget proposals being looked upon favourably by Congress.

Last year, the substantial increases for research and development requested by President Ford in his outgoing budget were received sceptically. Although making a considerable, and almost traditional, increase in the NIH budget, Congress reduced the request for the NSF, for example, from \$885 million—a figure which NSF Director Dr Richard Atkinson admits "we were not able to defend"—to \$863 million.

This year the OSTP has done the President's homework. And there will be both substantial evidence, and the carefully-considered support of departmental and agency heads, to back up demands in the face of Congressional cross-examination. So the signs are that 1979 will be a good year for scientists.

David Dickson

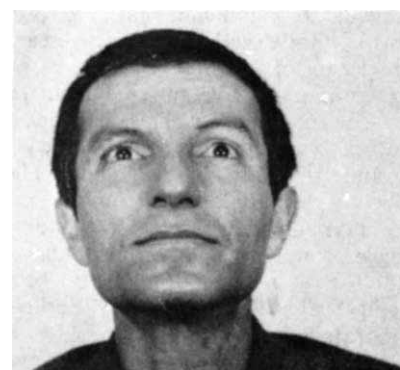
Federal funding for conduct of R and D by Agency (in millions of dollars)

Department or Agency	Obligations		% increase 1979 over 1978
	1978	1979	
Defense—military functions	11,709	12,740	8.8
Energy	4,231	4,245	9.3
NASA	3,877	4,193	8.2
HEW	3,137	3,258	3.9
NSF	754	829	10.0
Agriculture	626	632	1.0
Environmental Protection	351	358	2.0
Transportation	364	342	−6.0
Interior	366	340	−7.1
Commerce	288	316	9.7
All other*	587	637	8.5

*Includes the Department of Justice, Labour and State, The Veterans Administration, The Civil Service Commission, the Corps of Engineers, the Federal Trade Commission, the Tennessee Valley Authority, and the Library of Congress.

TOTAL	26.287	27,800	6.1
--------------	---------------	---------------	------------

Anti-relativist draws others into the whirlpool



Stefan Marinov

*O father Galileo, cunning one and wise,
Thy trial persisteth still even from age to age;
Moralist and philosopher try thee, the fool eke tries,
And everyone who counts himself a learned sage.*

*So wast thou then a coward, valourless, without honour,
Thyself knowing the truth, to spit on truth, deride,
Saving thy mortal frame, to fraud to sing "hosanna",
Before all men to trample thine honour and thy pride.*

*Holy lord of the spirit, my teacher wise and dear,
Is the common herd worth our torments and our blood.
Shout yourself hoarse—no sound will reach its blunted ears;
Throw your heart at its feet—onward it still doth plod.*

*So, doctors, I bow and swear: "There is no absolute space!
All I affirmed is lunacy—bring on your drugs apace!"*

RESTRICTIONS on scientific correspondence in the Soviet Union and its satellites have, over recent years, become familiar—the classic study of the problem being *The Medvedev Papers*. One of the most curious by-products of the system is the recent appearance, in Belgium, of an anti-relativistic tract with the lofty title *Eppur si Muove* and a preface signed by no less a person than A. D. Sakharov—presumably the dissident academician of that name. In fact, as the author of the book, Stefan Marinov, himself admits, Sakharov never wrote such a preface; Marinov claims, however, that Sakharov gave him permission to append his name to a preface written by Marinov on his behalf.

Although this may appear at first glance somewhat a trivial matter—that of a 'fringe scientist' trying to gain the backing of an eminent member of the orthodox community—the appearance of the preface could have considerable implications for Academician Sakharov. The various campaigns launched against him within the Soviet Union regularly imply that he has 'abandoned' or 'betrayed' science for 'so-called dissidence', and his apparent endorsement of a scientific theory which he himself does not hold, simply because its author was himself in trouble with the authorities in his own country, could add valuable fuel to this debate. The history of this curious preface is therefore worth looking into.

Stefan Marinov first made his appearance in the western media in the autumn of 1976, when large advertisements began to appear for a conference on 'Space and Time Absoluteness' the following May, on his initiative. A certain 'A. D. Sacharov' of Moscow was listed, variously, as Chairman or Patron of the conference. This surprising announcement led to considerable speculation, and a general consensus of opinion that it could not be Academician Sakharov who was meant. Even the difference in spelling

was cited to support this idea, by those who did not realise that the Russian name CAXAPOB would, in certain transliteration systems, be rendered as 'Sacharov'. In fact at the time of the announcement, Marinov and his western supporters were still trying to contact Academician Sakharov by telephone, to ask for his consent, and were approaching anyone (the present author included) whom they felt might be able to make such a contact.

Marinov's next attempt to contact Sakharov came the following spring. His *magnum opus*, refuting the theory of relativity and all associated physics, was ready for publication, and he wished Sakharov to provide a preface. Having still failed to contact Sakharov over the Varna Conference, Marinov wrote the preface himself, distributed copies to possible contacts with the request that they forward them to Sakharov, and added a covering letter which, in the manner of a student applying for an *exeat*, said that unless he heard to the contrary, he would assume that he had Sakharov's permission to proceed. In one version of the covering letter, he added a brief self-portrait. "As far as I know, I am the unique 'dissident' in my country (once in a prison, twice in a loony bin). I descend from an old family of intellectual communists, and I am a Marxist (I have even written a book on mathematical political economy—in Russian—and I have a translation in Serbo-Croatian). My opinions are most close to those of Roy Medvedev."

Marinov was soon to be back in the mental hospital for a third time. At the end of April 1977, telegrams signed 'Marinov' were sent to journalists and others who had any connection with the Varna conference, cancelling it on the grounds that an earthquake was expected. The immediate assumption, that Marinov had taken this means of cancelling an event which had no supporters, proved false. Marinov had been removed to hospital by the authorities,

who had then notified in his name all those on his address list. News of this reached the West in May, but journalists were earnestly requested by his friends not to publish, since this would endanger his life. In all events, once the critical dates of the planned conference were over, Marinov was released, and in late summer he was allowed to emigrate. He settled in Belgium.

In October, 1977, the news-magazine *Pourquoi Pas?* carried a massive article on Marinov, 'The Scientist who came in from the cold', with a reprint of the 'Sacharov' preface. This, allowing for translation and editorial omission, was identical with Marinov's own draft. Although it seemed highly unlikely that Sakharov would have lent his name, I decided to seek confirmation on this point. It is virtually impossible to get a letter through to Sakharov, and direct telephone contact is likewise a random matter with minute probability of success. Nevertheless, the message reached Sakharov by two channels, and two answers were received. One, via a physicist, ran 'Academician Sakharov knows of the book, but did not wish to be associated with it, as he does not agree with the theory!' The other, less formal message, was transmitted as 'Andrei Dmitrievich says: "The man's a nut-case (*psikh*), but I wouldn't want to condemn anyone to a mental hospital!"'.

At the end of November, Marinov turned up, uninvited, at the Science Session of the Venice Biennale. Asked about the preface, he maintained that a 'courier', described as 'an eminent physicist' and a 'young girl', had taken the book to Sakharov who received the courier, expressed sympathy for Marinov's plight, and agreed to 'think about' the matter of the preface. Sakharov is well known for his kindness and compassionate interest in all those in trouble; and he probably meant simply to give an expression of personal sympathy coupled with a polite refusal

to involve himself with Marinov's theories. Unfortunately, Marinov construed this as consent to have his signature added to the preface. Although a number of people entreated Marinov to withdraw it, he refused, saying that as it had appeared in *Pourquoi Pas?* it was now too late to do so. Moreover, he needed Sakharov's name to sell the book; unless he could sell 5,000 copies at \$20 each he could not get the money he needed to carry out the experiments described in it. (One presumes he meant 'replicate'.) A long and hysterical telex was dis-

patched to Sakharov c/o the Soviet Academy, and copies circulated among the Biennale journalists. Sakharov at that time was not even in Moscow; he and his wife were staging a sit-in in a Siberian labour camp where her nephew Edvard Kuznetsov, the dissident writer, had been refused his regular visit from the Sakharovs. At the time of writing, Marinov is still trying to get a message through to Sakharov.

Marinov's experiences in defence of his theories have undoubtedly made him only the more adamant in maintaining them. His poems imply that

his incarceration in the mental hospital was on account of his theories (see sonnet opposite). Clearly he is willing to take any means to promulgate them, even resorting to 'short cuts' when no answer is forthcoming. This is almost certainly not the first such occurrence in the long history of East European censorship—a number of very curious documents have reached the West from time to time. The whole episode is yet another illustration of the curious situations which can arise when governments restrict the freedom of scientific contact and correspondence.

Vera Rich

Polishing a tarnished image

LAST Wednesday the centre of Washington was brought to a standstill by a demonstration of angry farmers demanding "100% parity"—a price for their products that would give them the buying power of 65 years ago, when agricultural prices were at their peak. That same afternoon, a group of congressional employees was given a seminar on "government's role in scientific research" by a group of distinguished biomedical scientists, including three Nobel laureates—Arthur Kornberg, George Palade and James D. Watson—and the heads of some of America's leading biomedical research institutions and medical schools.

The style was different from that of the farmers, but the demand was very similar: a return to the levels of funding that basic research in the biological and medical sciences enjoyed in the relatively halcyon days of the late 1960s.

The case that the scientists presented during a well-organised two-day visit to Washington—which included private meetings with congressmen and members of the administration, as well as public hearings before the appropriations committees of both the Senate and the House—was straightforward. Basic research, they claimed, is grossly underfunded in comparison with applied research—it is in a chronic state of instability and lacks the means of training a new generation of scientists.

"We are here to draw to the attention of our legislators the importance of basic biological research in the solution of major elements of our nation's health care problems," Dr Mahlon Hoagland, President of the Worcester Foundation for Experimental Biology, and a major organiser of the Washington visit, told Senator Thomas Eagleton's appropriations subcommittee on the budget of the Department of Health, Education and Welfare.

On the surface, the argument was

about money; the report in the *New York Times* carried the not unfamiliar headline "Scientists plunge into lobbying for more medical research aid". And the scientists presented a carefully-quantified list of grievances.

For an example they claimed that there has been an 18% drop in the total amount of federal funds spent on basic research since 1967, and a reduction of 17% in support for scientist-initiated grants awarded by the National Institutes of Health (NIH) between 1967 and 1975.

Between 1967 and 1977 there was a decrease in the proportion of grants funded to grant applications submitted from 53% to 33%, they told the subcommittee. And the scientists also pointed out that there has been a decrease in funds for training young scientists from 18% of the NIH's extra mural budget in 1967 to 6.8% in 1976.

The demands, too, were specific. The group said that it wanted the NIH to be provided with an across-the-board increase in funding of 10% in the fiscal year 1979 to compensate for the effects of inflation, and a return to 1967 levels in both scientist-initiated grants for basic research (then 61% of the NIH external budget) and the biomedical research support grants system (then 7%).

In addition, they requested an extra \$100 million a year for five years (an increase of almost 50% over the current budget) to be added to the budget of the National Institutes of General Medical Science, the NIH's basic research institute through which many biomedical research activities in universities and medical schools are funded.

Yet the visit to Washington was not only—or indeed primarily—about money. Indeed on purely statistical grounds, the case that the scientists presented lay open to criticism. It was pointed out, for example, that by taking the 1967 figure as a bench-mark, a

year in which research funding is generally reckoned to have reached the peak of the 1960s expansion, figures for subsequent years appear particularly—and perhaps artificially—bad. And figures presented purely as percentages obscure the almost 300% increase in total funding for NIH.

Furthermore both NIH and President Carter's Office of Science and Technology Policy have shown themselves to be aware of the current problems facing the basic research community. After what everyone agrees was a bleak period between 1967 and 1972, funds for basic research have been picking up, and will continue to do so if Congress accepts the suggested increases in President Carter's budget proposals presented this week.

But behind the dispute over financial resources lies a deeper issue of concern to the scientific community, the public image of science, and in particular of basic science on which Congress's willingness to provide additional funds ultimately stands.

In recent years, just as the debate over the implications of the Rothschild Report in Britain have reflected growing demands for the "relevance" of medical science, so similar tendencies in the US have given rise to what has been called the "disease of the month" mentality with a philosophy that medical science should be primarily directed towards curing, rather than understanding a disease.

In this climate, as funds have come pouring in for research into disease-related programmes such as cancer and heart disease, resulting in the total NIH budget increasing from about \$1,000 million to over \$2,500 million in seven years, so basic research has—in relative terms—lagged behind, and the process of scientific discovery has, it is claimed, been both distorted and delayed.

In the eyes of many basic scientists the villain of the piece is the so-called "war against cancer" launched in 1971

with President Nixon's much-heralded National Cancer Act, and responsible for an increase in cancer research funding from \$180 million in 1970 to almost \$900 million in 1978.

No one is claiming that the money has been entirely wasted. Much good science has been carried out under the cancer programme, and despite occasional widely-publicised lapses, few are prepared publicly to identify specific projects which they feel should not have been funded.

But various factors have led to a cooling off in Congress' initial enthusiasm for a massively-financed cancer research programme, and to increasing demands for a visible pay-off from its investment. These factors include data showing that, in spite of all the research and clinical advances, deaths from cancer continue to increase, and the growing evidence that many cancers are due to environmental causes, accessible to preventive rather than curative techniques.

Congress' frustration at the lack of tangible results reinforces the view of those scientists who criticised the whole "target-oriented approach", to biomedical research funding from the beginning. The fear, however, is that a failure of strategy could result in a general disillusionment with the whole research enterprise, affecting both its basic and applied aspects.

Dr Arthur Kornberg, for example, professor of biochemistry at Stanford University, claims that it is very rare for a crash programme of biomedical research to succeed in its objective, but warns that failure also carries its price. "Progress in medicine rests on fundamental advances. You harm people by trying to do things prematurely—you destroy your credibility, and the whole of science suffers. We have seen this happen in recent years, and people are discouraged from entering a field which has come to be regarded as bad science."

Dr Kornberg criticises the extent to which research workers are increasingly required to keep their eyes on a fixed target. "At present we have to boot-leg under various guises if we want to carry out fundamental research. This has both the spirit and the content of scientific investigation."

The task therefore facing the biomedical community, confronted by an apparent failure to come up with the goods that congress—many feel unjustly—has demanded, is how to make it respectable to be seen giving money to basic research in the biological and medical sciences. Criticism of the cancer programme, for example, will be to no purpose if the net result is an overall reduction in NIH research funds, rather than a redirection of



Dr Arthur Kornberg

funds towards basic research.

Furthermore Congress has no direct control over the way that the directors of the various NIH institutes distribute the funds allocated to them. A suggestion that each institute be required to spend at least 30% of its budget on basic research would be "a mistake" according to Dr Donald Frederickson, Director of NIH, who points out that most institutes do this already.

The best that scientists can hope for is some recommendation from Congress that it would like to see the various institutes upgrade basic research, a suggestion to be checked later against actual performance. "I feel that some of the institutes have tended to overemphasise target research at the expense of basic research because that was what Congress wanted. But Congress might now take the opportunity to communicate its belief in the importance of basic research," Dr Seymour Kety, Professor of Psychiatry at Harvard University and a past scientific director of the National Institute for Mental Health told the Senate Appropriations Subcommittee.

So far, the response of several key Congressmen has been relatively favourable. Sensing a growing disillusionment with the target-oriented approach to research funding, they are prepared to back the case that the long-term solutions lie in supporting basic research.

Mr Eagleton, for example, told the scientists that his subcommittee had "become aware of the pitfalls brought on by the proliferations of targeted research programmes." And Representative Paul D. Rogers, Chairman of the House Subcommittee on Health and the Environment, has promised to introduce a research training grants bill that would provide an extra \$220 million in 1979.

But the case is far from conceded. Cancer research still has powerful supporters who, while accepting that it is

the quality rather than the quantity of research that matters, claim that this means more rather than less money. And there remains the widespread feeling in many quarters—frequently voiced, for example, by Senator Edward Kennedy—that medical scientists should be required to produce results of visible social usefulness to justify the large investment of public funds that they receive.

Thus in spite of a significant increase in federal support for basic research in the proposed budget for 1979, basic biomedical science is unlikely to achieve its "100% parity" with the boom years of the late 1960s. At least not in the near future. The most that scientists can hope for is that they can encourage the pendulum to swing a little faster in their direction.

David Dickson

Lederberg named President of Rockefeller University

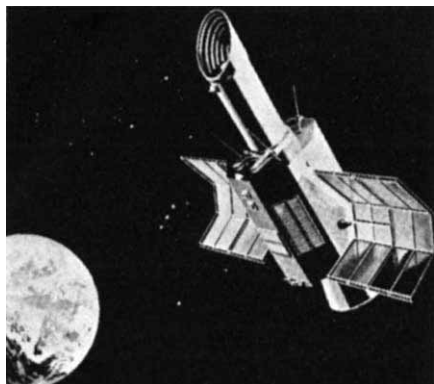
Professor Joshua Lederberg, Chairman of the Department of Genetics at the Stanford University School of Medicine, and Nobel prize winner in 1958 for his work on the organisation of genetic material in bacteria, has been elected President of the Rockefeller University in New York. Professor Lederberg will take up his appointment from 1 July and succeed Dr Frederick Seitz.

UK science budget announced for 1978/9

Spending on science through the UK research councils will be up by more than 2% in real terms in 1978/9, claim the Advisory Board for the Research Councils (ABRC) following agreement between the Board and the Secretary of State for Education and Science on expenditure levels. The Science Budget, which funds the research councils, Natural History Museum and Royal Society, is £256 million at 1977 prices; the Science Research Council get £139 million of it. The figures include two increases recently authorised, one for a recurrent £4 million following a slight easing of the economic situation, the second for a once-off £4.5 million to stimulate the construction industry.

The distribution of the budget still reflects ABRC's intention gradually to redeploy resources away from big science but there is now sufficient leeway with the extra money for one person at SRC headquarters to describe the mood there as "reasonable satisfaction, even modified rapture".

International Ultraviolet Explorer to be launched this week



DURING the course of evolution some animals have developed eyes on stalks as a rather bizarre adaptation to their environment. Astronomers could well do the same to escape the disturbing effects of the earth's atmosphere but evolution is a slow process and, luckily, there are more acceptable solutions. For many years now, astronomers working at wavelengths that are absorbed by the atmosphere have been sending probes into space, but the available techniques have not been entirely satisfactory. Rockets and balloons provide only fleeting glimpses and orbiting satellites not continually in communication with the ground must be automated to collect data blindly for later transmission. A solution is required which enables the astronomer continuously to control and monitor the performance of a telescope that remains above the atmosphere for long periods of time.

That solution is provided for the ultraviolet (UV) wavelength region by the International Ultraviolet Explorer (IUE) satellite scheduled for launch into geosynchronous orbit on 26 January. The satellite will be continuously visible from a control station on the ground and can be used just as if it were a ground-based UV telescope not suffering from the effects of the atmosphere.

In the ultraviolet region those effects are disastrous. The earth's atmosphere becomes increasingly opaque at wavelengths below 4,000 Å and is effectively 'black' beyond 3,000 Å. Balloons can observe over most of the mid-ultraviolet range from 3,000 Å to 2,000 Å by ascending above the bulk of the ozone layer to heights of 40 km, but for the far ultraviolet below 2,000 Å it is necessary to go much higher. The planned orbit of IUE varies from 25,000 km to 46,000 km altitude.

There are sound reasons for wanting information at ultraviolet wavelengths. The sun is brighter at visible wavelengths than in the far ultraviolet but this is not true for many other stars. As the temperature of a source

increases, the emitted radiation appears at shorter wavelengths and at 20,000 K, a typical surface temperature of a newly-formed star, almost all of the light is at wavelengths that cannot penetrate the earth's atmosphere.

The cooler regions between the stars are also important at UV wavelengths. Interstellar gas contains about 10% of the material in our galaxy and can be detected only by the characteristic absorption lines it imposes on the spectra of objects observed through it. The gas constituents are in their lowest energy state because of the low densities and temperatures in space and so it is necessary to observe at wavelengths where the atoms and molecules in their ground states can absorb light.

The common elements in interstellar gas are hydrogen, helium, oxygen, carbon, nitrogen and silicon. They all have their strongest absorption lines below 3,000 Å, with only the less abundant sodium and calcium appearing in the visible region. Therefore the major constituents of interstellar gas clouds can only be observed by measurements in the UV. If the gas clouds are moving, then the absorption lines will be Doppler shifted in wavelength and high resolution measurements of the spectra allow their velocity to be calculated.

Curiously, UV emission from some of the stranger objects in the sky has been studied from the ground. Many of the immensely distant quasars are receding so fast that their UV light is Doppler shifted into the visible region; they are said to have large redshifts. These cannot be compared directly with low redshift quasars whose UV emission is shifted less and still remains in the band absorbed by the atmosphere. Direct comparison is essential to find out if quasars are at cosmological distances, and the Lyman alpha line is crucial as it may provide a measure of the intrinsic brightness of quasars. So far, quasars have been disappointing cosmological objects simply because they could not be easily observed in the ultraviolet. But a recent UV observation of a low redshift quasar has imposed important constraints on cosmological models. Those measurements were made from a rocket flight and lasted only four minutes—IUE can observe for 24 hours a day.

The present knowledge of the appearance of the universe in the UV comes from balloon and rocket flights and from experiments on board the American Orbiting Astronomical Observatory satellites OAO-2 and OAO-3 (renamed Copernicus) the Dutch ANS satellite and the European TD-1 satellite. IUE is dedicated solely

to UV observations and promises an enormous improvement on the earlier, largely survey, experiments. The IUE project is itself a forerunner for the Space Telescope planned to be launched from the Space Shuttle in 1983.

There are two broad aims of IUE and these impose conflicting requirements on the instrumentation. The measurement of chemical abundances, and the study of gas motions, require a detailed study of line profiles with very high spectral resolution. But IUE should be able to detect faint sources making necessary good light collection efficiency rather than resolving power. The ability to observe for long periods also means that variable phenomena can be observed.

To satisfy these requirements, IUE carries a 1.3 m long, 45-cm diameter aperture, Cassegrain telescope which will be used exclusively for spectroscopy. No direct imaging will be made in the UV although a visible-light picture will be used for target identification. Light from the telescope is passed through two spectrographs, each containing a high dispersion echelle grating in series with a low dispersion grating; two spectrographs are needed to achieve resolutions of between 0.1 Å and 0.2 Å over the full wavelength range of 1,150 Å to 3,200 Å. For the low resolution mode, the echelle is simply bypassed, producing a spectrograph that can be used to record stars as faint as 14th magnitude with a 6 Å spectral resolution.

IUE is a joint venture between the US National Aeronautics and Space Administration (NASA), the European Space Agency (ESA), and the UK Science Research Council (SRC). The major contributor is NASA. ESA is providing the solar panels and one of the two ground stations and the SRC is supplying the UV sensitive television cameras used for spectra detectors. The observing time on IUE will be shared on a daily basis with 16 hours going to NASA, and 4 hours each to ESA and the SRC who have agreed to take 8 hour periods on alternate days.

As the telescope is in geosynchronous orbit it will be in continuous contact with the NASA ground station at the Goddard Space Flight Center and will be in view at least 10 hours a day from the ESA ground station at Villafranca near Madrid. This arrangement allows real-time operation of the telescope and reduces the training time necessary for observers. As a result, the telescope can be opened up for international use and guest observers are encouraged to use IUE in a similar fashion to ground based telescopes.

Stuart Sharrock

correspondence

Conserving uranium

SIR,—We read with interest the recent article by John Davies on this topic (1 December, page 376). However, it does contain a number of scientific errors and misleading statements which we believe it is our duty to point out.

He is incorrect in stating that the difference between the $^{238}\text{U} \rightarrow ^{239}\text{Pu}$ and the $^{232}\text{Th} \rightarrow ^{233}\text{U}$ reactions is that the first needs fast neutrons while the second can use slow ones. In fact the cross-sections for the two reactions are almost identical at all energies and therefore they have very similar breeding characteristics. The reaction $^{238}\text{U} \rightarrow ^{239}\text{Pu}$ occurs to a very considerable extent in thermal reactors. In fact about one third of the energy output of a thermal reactor arises from the ^{239}Pu produced within it and then burned *in situ*. Thus in the UK we already get between 4 and 5% of our electricity from the fission of plutonium.

The control of both thermal and fast reactors is based on the delayed neutron contribution and this is on a time scale of seconds rather than micro-seconds. The only mechanisms for a sudden increase in reactivity are loss of coolant or fuel movement, neither of which could occur faster than the control rods could move.

Turning to the $^{233}\text{U}/^{232}\text{Th}$ fuel cycle we would first challenge the statement that ^{233}U is automatically safeguarded against illegal diversion by the intense γ -rays coming from the simultaneously produced ^{232}U . This isotope does not emit γ -rays. It decays with a 70 year half-life by α -emission to ^{228}Th which does emit the γ -rays to which Dr Davies refers. Thus only after a considerable period does the γ -activity build up to a sufficient level to afford intrinsic protection.

With regard to the waste disposal problem the thorium cycle may give rise to fewer problems from the production of long lived higher activities but on the other hand much more Pu^{238} and ^{234}U is produced. The latter isotope decays to the very hazardous ^{226}Ra .

Accelerators are replacing reactors for some forms of neutron research because they can produce neutrons in short pulses and can give higher peak fluxes than a reactor. However, the mean flux from such systems up to present is lower by orders of magnitude than in research reactors.

The use of accelerators for breeding significant quantities of fuel demands machines giving outputs two or three orders of magnitude greater than the most powerful machines built so far. The breeder target is nothing less than a very major reactor in its own right. Studies of such a system at Brookhaven suggest a cost/gm of ^{239}Pu three to five times greater than that of equivalent enriched ^{235}U . The hazards of such a system could certainly not be less than that of a major reactor system.

We do not wish to give the impression that we are in any way antagonistic to the thorium cycle or to accelerator breeding. We wish merely to warn against the temptation of assuming, like the grass being greener on the other side of the fence, that problems of an alternative technology are necessarily less than those of a technology upon which one has embarked.

Yours faithfully,

A. T. G. FERGUSON

I. M. BLAIR

AERE Harwell

Transposed sentence

SIR,—It has been brought to the attention of the editors of volume 23 of *Biographical Memoirs of Fellows of the Royal Society* that certain sentences in the memoir on Professor C. H. Waddington may be read as derogatory of the work of Dr Ruth Clayton and her colleagues. This was in no way the intention of the author, but arose from an unfortunate transposition of sentences in an early draft. The author therefore wishes to amend the last six lines of page 583 to read:

"... work that Waddington had originally envisaged. A notable exception which fitted into Waddington's original concept was the work on the development of the avian lens in the group led by Mrs Ruth Clayton. The laboratory became well known for the DNA-RNA investigations which earned a world reputation for at least two of the leading workers."

We must unreservedly withdraw all inferences to the effect that Mrs Clayton's work and that of her group is anything other than of the highest class and of a deserved worldwide reputation and we whole-heartedly apologise to Mrs Clayton for the embarrassment caused to her and her

group by the unfortunate misplacing of the sentences in the article.

Yours faithfully

R. W. J. KEAY

The Royal Society,
London

Anti-Müllerian hormone

SIR,—In a recent issue of *Nature* the authors D. Tran, N. Meusy-Dessolle and N. Josso report on a testicular factor distinct from testosterone which mediates regression of Müllerian ducts in male foetuses (*Nature* 269, 411–412; 1977). Apart from the heading of their article which read "Anti-Müllerian hormone, etc." (and which caused underlining and exclamation marks by some of my merry colleagues) I do not consider it appropriate to use laboratory-born artificial words such as 'testicular anti-Müllerian activity', 'Müllerian regression' and others in a scientific communication. I wonder how this obvious jargon escaped your referee's attention*. Besides there is no 'Müllerian regression' but a regression of the outline, the diameter or so of the Müllerian duct; likewise there is neither a 'testicular anti-Müllerian activity' nor an 'anti-Müllerian hormone'. I am sure that Johannes Peter Müller (1800–1858) would agree.

Yours faithfully,

GERHARD H. MÜLLER

Universität des Saarlandes
Saarbrücken, West Germany

*It didn't, but the authors claimed that since the name had been used in all their recent publications, a change now would cause confusion.—ED.

Essential relativity

SIR,—In connection with Paul Davies' kind review of the second edition of my book *Essential Relativity* (*Nature* 22 September 1977) I wonder if you would allow me to correct one slight error. The coordinates of accelerated observers in Minkowski space, which have recently been used in the study of black holes, are elaborated on, albeit in a later section on Kruskal space. Indeed, this was one of the more significant additions to the second edition.

Yours faithfully,

WOLFGANG RINDLER

The University of Texas at Dallas

news and views

The end of the expanding Earth hypothesis?

from Peter J. Smith

ON the face of it, this is a bad day for the handful of people who support the idea of an expanding Earth, a good day for the few who oppose it, and certainly no less than an interesting day for the vast majority of Earth scientists who have been content merely to observe the progress of the expansion debate for anything up to 20 years. For on page 316 of this issue of *Nature*, McElhinny *et al.* offer the most convincing proof yet that, despite some obvious attractions, Earth expansion is nothing more nor less than a scientific blind alley. Specifically, they use the best available palaeomagnetic data to show that over the past 400 Myr the Earth's radius cannot have increased by more than 0.8%, a figure sufficiently small to exclude the very low rate of expansion (0.6 mm yr^{-1}) proposed by Wesson (*Q. Jl R. Astr. Soc.* **14**, 9; 1973) as well as the much higher rates favoured by, among others, Carey (in *Continental Drift—A Symposium*, University of Tasmania, 1958).

So that is the end of the Earth expansion hypothesis. Or is it? Before passing the death sentence it is perhaps worth making two points which may be interpreted as reasonable caution or unreasonable scepticism, depending upon one's point of view. First, history has shown that where the Earth sciences are concerned it is often dangerous to express absolute certainty. The continental drift saga alone is sufficient indication that even the most solid of foundations may crumble, especially if they are given a kick by a passing stranger. In the case of continental drift the stranger was a meteorologist, but for Earth expansion it could well be the first physicist or astronomer to delineate the historical variations, if any, in the gravitational constant. Earth scientists have been conscious of the possibility of Earth expansion ever since Dirac (*Proc. R. Soc. A* **165**, 199; 1938) suggested that

the gravitational constant may be slowly decreasing; but although convincing proof of such a decrease has never yet been produced there is still an outside chance that it will be. In that case, of course, expansion of the Earth would have to be accepted; and however hard it may be for them, Earth scientists would have no option but to re-examine their supposedly fool-proof analyses.

But if that is a possible problem for the future, there is also the more immediate question of the validity of palaeomagnetic techniques for determining the Earth's palaeoradii. As McElhinny and his colleagues point out, Egved (*Geofis. Pura Appl.* **45**, 115; 1960) was the inspiration for the use of palaeomagnetic data as a test for Earth expansion. His first proposal was limited to palaeomagnetic sampling sites on the same palaeomagnetic meridian, but he later (*Nature* **190**, 1097; 1961) suggested a less restrictive technique applicable to palaeomagnetic data generally. In the much improved version developed by Ward (*Geophys. J.* **8**, 217; 1963; **10**, 445; 1966) this 'minimum scatter method' involves the determination, for a single landmass, of palaeomagnetic pole positions for a series of assumed palaeoradius values. The 'correct' palaeoradius is then taken as the one corresponding to the minimum scatter of poles. Unfortunately, in the hands of Ward and others the method achieved only limited success, for at that time palaeomagnetic data were neither accurate nor extensive enough to enable any but the greatest of suggested palaeoradius changes to be detected.

That position has now changed, however, and McElhinny and his coworkers have been able to revive Ward's method in conjunction with palaeomagnetic data capable of resolving the smallest changes in the Earth's radius ever proposed. But irrespective of the quality of the raw data, is Ward's method valid in the first place? McElhinny *et al.* evidently believe that it is; but Carey (*The Expanding Earth*, Elsevier, 1976) has already argued at

great length that it is not. He concludes that the minimum scatter of poles will always occur when the assumed palaeoradius of the Earth is about equal to the present radius; so however good the basic data and however much the Earth's radius has really changed, the application of Ward's method will inevitably lead to the conclusion that the Earth has not expanded at all.

Ward assumes that although the shape of the continents must change slightly as the curvature of the Earth's surface changes, they remain constant in size and the distances and angles used in interpreting the palaeomagnetic data in terms of the palaeoradius also remain constant, when measured from a 'central point'. Ward took as the 'central point' the mean position of the rock units studied; Van Hiltten (*Tectonophysics* **5**, 191; 1967) later took the centroid of the continent; McElhinny *et al.* (who refer to the "continent keeping the same physical dimension") use both the average site location and the "approximate centre of the continental block", discovering in the process that both lead to very similar values of palaeoradius.

Carey contends, however, in a long and involved geometrical argument that during expansion of the Earth a continent will deform in such a way and to such an extent as to make nonsense of any simplified model used in the determination of palaeoradius from palaeomagnetic data. If he is completely right, palaeomagnetism presumably has no role in detecting any possible increase in the terrestrial radius. If he is right in principle but not practice (perhaps because the continental deformation is too small to invalidate Ward's model), palaeomagnetic methods will be acceptable as long as the basic data are good enough. If he is wrong, there is nothing to worry about. The fact is, however, that whether Carey is right or wrong his criticism exists and has apparently never yet been refuted explicitly. Until someone chooses to do so, there must be lingering doubt in the minds of disinterested observers. □

Peter J. Smith is a Reader in the Department of Earth Sciences at the Open University.

Progress on slime mould morphogens

from Peter C. Newell

ALTHOUGH the cellular slime mould *Dictyostelium discoideum* has yielded a great deal of information about the intercellular signalling involved during the organism's aggregation phase (the phase when the starving amoebae swarm together to form a multicellular body) there has always been the possibility that it would be just as difficult to get one's hands on the intercellular regulatory substances or 'morphogens' that are used during the later stages of morphogenetic development as it has been in more complex embryonic systems. Recent reports from two groups working with *D. discoideum* however, give strong encouragement for a more optimistic point of view.

Three classes of morphogen may be envisaged in this organism. For example, one might be concerned with controlling which of alternative pathways of development are taken, a second with inducing development in general without regard to the cell types formed, and a third with inducing the differentiation of specific cell types. In the first class is the substance that can switch developing aggregates of cells from the sorocarp (stalk and spore)-forming programme to the alternative 'slug' programme. During the slug programme stalk and spore development is held in abeyance while the organism slowly moves over the substratum in search of a more suitable place to develop. The slug is morphologically and behaviourally distinct from the alternative forms of development and is apparently formed in response to an 'awareness' by the developing aggregate of an overcrowded environment. In a recent paper by Schindler and Susman (*J. molec. Biol.* **116**, 161; 1977) the substance that induces formation of this migratory form is convincingly identified as NH_3 . This metabolite is formed in large quantity by developing aggregates as the starving cells rely on the breakdown of endogenous protein for a major part of their energy requirements. The evidence that NH_3 is the regulating agent comes from the finding that, not only does the naturally occurring agent show many of the chemical and physical properties of NH_3 but also that it is mimicked by ammonium carbonate (at a pH above 7.0) when added during the appropriate 'decision-making' period. In addition the slug stage can be omitted entirely if the NH_3 -absorbing system of glutamate dehydrogenase, NADH and α -ketoglutarate is added to the young aggregates.

The NH_3 seems, therefore, to be acting as a simple, metabolically produced morphogen that regulates the developmental pathway taken by the organism.

In the second class of morphogen is the inductive signal for differentiation into spores and stalk cells. From work originally performed by John Bonner (*Proc. natn. Acad. Sci. U.S.A.* **65**, 110; 1970) the regulatory substance at this level of control seems to be cyclic AMP. Cyclic AMP causes cells to transform into typical stalk cells without ever aggregating in the normal way. It is also the chemoattractive substance by which starving *D. discoideum* aggregate ready for development. However, in addition to this action there is evidence that pulses of cyclic AMP released during the aggregation process (rather than a continuous signal) are able to switch on some of the very early events of differentiation (Gerisch *et al.* *Nature* **255**, 547; 1975). Moreover, the genus *Polysphondylium*, in which cyclic AMP, although secreted at late stages of aggregation, is not a chemoattractant (this role is taken by a polypeptide instead) is also induced to form stalk cells by added cyclic AMP, (Francis *Nature* **258**, 763; 1975). The two actions of chemotaxis and stalk cell induction are therefore clearly distinct. Until very recently one was tempted to deduce from these observations that cyclic AMP has a specific effect on stalk cell induction. That this is not so is ably shown in the recent paper of Kay, Garrod and Tilley (*Nature* **271**, 58; 1978) who took, as a measure of spore induction, the number of cells induced to form the 'prespore vacuoles' (PSVs) that are characteristically formed during the course of spore maturation, rather than looking for mature refractile spores as previously. Using this criterion of spore pathway induction they found that both spore and stalk cell development was initiated by added cyclic AMP. Spore differentiation differed from that of stalk cells in that the process could not be completed to maturity under these conditions. Thus although cyclic AMP can induce differentiation, it must be a general inducer for the slime mould cells.

Important progress has also recently been made in demonstrating the existence and some of the properties of the specific cell-type inducers that can be considered as the third class of morphogen. The evidence for the idea that cyclic AMP on its own was not sufficient for stalk cell induction came from the work of Town, Gross and Kay

(*Nature* **262**, 7; 1976) who found that the proportion of cells forming stalk cells in monolayers in the presence of cyclic AMP was dependent on the cell population density. The farther apart the cells were separated the fewer stalk cells were formed. Using Cellophane membranes to separate cell layers at different densities Town *et al.* were also able to show that cells at high density on one side of the membrane were able very efficiently to help cells at low density on the other side to develop into stalk cells in the presence of cyclic AMP. The diffusible factor responsible is currently being characterised and preliminary results show that it is a small glycopeptide of 1,000 to 2,000 molecular weight whose biological activity is dependent on sialic acid, N-acetylglactosamine and L-fucose that are present in its carbohydrate moiety (C. Town, personal communication). Using a similar technique but monitoring PSV formation to identify cells in the spore-forming pathway, Kay *et al.* described in their recent paper that although prespore cell induction is also cell density dependent (suggesting that, again, an endogenous factor is involved) the helping cells are only properly effective if present on the same side of the membrane as the cells being induced. In contrast to the stalk-specific factor the spore induction factor is not, therefore, freely diffusible through such membranes and can only act when cells are in very close proximity to each other.

The relationship between the three classes of morphogen is under active investigation. The area in which most progress has been made so far is that between NH_3 and cyclic AMP. In a model recently proposed by Sussman, Schindler and Kim (In *Development*



A hundred years ago

It is gratifying to know that at last Cleopatra's Needle has safely reached the Thames. It is proposed to moor the ingeniously constructed vessel containing the obelisk at a convenient part of the Thames embankment for some days, to enable the public to inspect it.

From *Nature* **17**, 24 January, 251; 1878.

Peter C. Newell is a Fellow and Tutor in Biochemistry of St Peter's College and a lecturer in the department of Biochemistry, University of Oxford.

and Differentiation in the Cellular Slime Moulds (Eds Cappuccinelli, P., Ashworth, J.) 31, (Elsevier/North Holland, 1977)) it was suggested that NH_3 inhibited spore and stalk cell differentiation by inhibiting cyclic AMP production. Recent work of Schindler and Sussman (*Biochem. biophys. Res. Commun.* **79**, 611; 1977) showing that NH_3 (but not NH_4^+) at physiological levels dramatically decreases cyclic AMP synthesis or release in starving cell suspensions and the work of Thadani, Pan and Bonner (*Expl. Cell Res.* **108**, 75; 1977) showing an inhibitory effect of NH_3 on cyclic AMP release by cells of the species *D. mucoroides*, suggest that this conjecture may indeed be soundly based. Whether the effect is indirect or whether NH_3 acts directly on the membrane-bound adenylyl cyclase remains to be resolved. $\square$

Energetic oxygen ions in the magnetosphere

from D. T. Young

MAGNETIC storms are vast disturbances of the Earth's magnetosphere dissipating energy at the rate of 10^{11} – 10^{12} watts over several days. Their connection to disturbances on the Sun was realised more than 100 years ago and it is now clear that solar weather is carried to the Earth by the flow of solar wind plasma past the magnetosphere. Disturbances in the solar wind step up the vast magnetospheric dynamo thereby providing the storm's energy. But basically magnetic storms require the acceleration and transport of large amounts of hot plasma (0.1–50 keV) and it is something of an embarrassment that no one really knows where all this plasma comes from. Does it originate as cold ($\lesssim 0.1$ eV) ionospheric plasma or as relatively hot solar wind (electrons ~ 10 eV, ions ~ 1 keV per nucleon)? Given what is known about magnetic storms and their relation to the solar wind the most reasonable assumption to make is that the hot plasma also originates in the solar wind. That this is not entirely the case has been demonstrated convincingly by the recent observation during three magnetic storms of roughly equal amounts of hot O^+ and H^+ ions (0.5–16 keV) in the magnetospheric ring current (Johnson *et al.* *Geophys. Res. Lett.* **4**, 403; 1977).

Centred about the equatorial plane, the ring current is an enormous electric

circuit whose current is carried primarily by ions trapped in the geomagnetic field. Although the 'ring' occupies much the same region of space as the Van Allen radiation 'belts' the latter are much less dense and contribute very little net current. Magnetic storms are manifestations of the growth and decay of the ring current as hot plasma is injected into it early in the storm and is then gradually lost, primarily by charge exchange with neutral hydrogen and by scattering into the atmosphere. During storms the plasma particles penetrate deeper into the geomagnetic field and the quiescent ring current (normally centred at altitudes of $\approx 30,000$ km) shrinks down to altitudes of $\approx 10,000$ km. Although Johnson and his colleagues at the Lockheed Palo Alto Research Laboratory obtained data at somewhat lower altitudes ($\lesssim 8,000$ km), the bounce motion of the ions along magnetic lines of force enables them to sample the actual current carriers.

Together with the Lockheed group's earlier report of O^+ and H^+ ions streaming upwards out of the ionosphere (Shelley *et al.* *Geophys. Res. Lett.* **3**, 654; 1976) their present results have far reaching consequences for models of the aurora and storm-time ring current. Most of these rely heavily on the hypothesis that solar wind plasma enters the magnetosphere by way of the magnetotail on the night side of the Earth. Plasma acceleration and transport deeper into the magnetosphere occurs during the 'collapse' of part of the magnetotail and the accompanying conversion of stored magnetic energy into particle kinetic energy. As the heated plasma is driven deeper into the nightside magnetosphere, the more easily scattered electrons travel down magnetic field lines, bombarding the atmosphere and producing the aurora. Gradients in the geomagnetic field cause the relatively more stable ions to continue drifting in a westerly direction eventually forming the ring current. This explains why what was often called the 'proton' ring current was expected to have solar wind like composition and why it is so disconcerting to find terrestrial ions mixed in apparently large proportion. How the generally accepted hypothesis of plasma transport and acceleration from the magnetotail earthwards will be reconciled with the Lockheed group's observation of ionospheric plasma accelerated in more or less the opposite direction remains to be seen.

Johnson *et al.* obtained their data with one of a new class of particle detectors, the energetic ion mass spectrometer. Only in the past few years has it been technically feasible to build mass spectrometers with the required wide energy and mass ranges (typically

≈ 0 –15 keV per charge and 1 to >30 a.m.u. respectively) within the rigid constraints of low weight and stray magnetic fields demanded of satellite instrumentation. Perhaps the most sophisticated of the new spectrometers was first flown on the ESA GEOS satellite (Balsiger *et al.* *Space Sci. Instrum.* **2**, 499; 1976) while the more recently launched ISEE-A carries a nearly identical spectrometer built by the Lockheed and GEOS groups together. To appreciate that these instruments are indeed a step forward one need only consider that before 1972, plasma detectors operating between 10 eV and 50 keV were either electrostatic analysers or solid state devices used only to measure particle energy and classify their charge as positive or negative. Since solar wind plasma is known to be $\approx 95\%$ protons the energy-only detectors abetted the hypothesis that the hot magnetospheric plasma originates in the solar wind. With this assumption to go on it was not necessary to measure anything but 'protons'.

Adding to the confusion over magnetospheric plasma origins brought on by the Lockheed results are observations made with the GEOS mass spectrometer of O^{2+} and $^4\text{He}^{2+}$ inferred to be of terrestrial origin (Young *et al.* *Geophys. Res. Lett.* **4**, 561; 1977). We conclude that the ions are produced at altitudes $\gtrsim 20,000$ km (that is, not the ionosphere) and are accelerated during storms to energies as high as 1 keV. This result is disconcerting because it was long thought that the simplest test for solar wind plasma in the magnetosphere would be the ratio of $^4\text{He}^{2+}$ to protons. This is about 0.04 in the solar wind and, before GEOS, the terrestrial value at keV energies was thought to be orders of magnitude smaller. The $^4\text{He}^{2+}$ ion is even further compromised as a tracer because it is converted rather quickly by charge exchange to $^4\text{He}^+$. As tracers one is apparently left only with more exotic (and difficult to measure) solar wind species such as O^{6+} and ^3He . The latter has in fact been detected inside the magnetosphere by the foil collection technique.

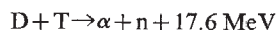
Mass spectrometers have been something of an ugly duckling in magnetospheric research. Initially too heavy, magnetically dirty and ill suited for hot plasma measurements, they have come of age and are now invited to all the best satellite projects. The GEOS-ISEE type spectrometer will be flown on GEOS2 scheduled for launch this year, and is included on NASA's next magnetospheric research satellite the Dynamics Explorer. It therefore seems likely that energetic ion mass spectrometers will have the opportunity to supply a few answers to the many questions they have been raising lately.

D. T. Young is at the Physikalisches Institut, University of Bern.

Thermonuclear reactions confirmed in laser-driven fusion

from Derek Beynon

THE central concept in inertial-confinement fusion is the implosive compression of deuterium-tritium fuel to densities and temperatures where a significant number of thermonuclear reactions will occur. Throughout the world a number of experimental programmes exist to investigate the capacity of powerful lasers, relativistic electron beams and, latterly, heavy-ion beams to achieve this mode of fusion. The ideal compression should occur isentropically in an ablatively imploded pellet of cryogenically solid deuterium-tritium mixture with a final density compression of about 10^4 in order to achieve an energy 'break even' where the resulting fusion energy equals the laser input energy (*Nature* **239**, 139; 1972). Because of the power output limitations of the present generation of large neodymium glass or CO_2 lasers, which give up to 10^{15} – 10^{16} W cm^{-2} on target, experiments have been limited to 'exploding pusher' targets which are glass microshells, typically 90 μm diameter, with 1 μm wall thickness, containing an equimolar deuterium-tritium gas mixture of density 2–2.5 mg cm^{-3} . Producing a non-isentropic compression, the whole of the glass shell heats rapidly before it moves to push the enclosed gas to about normal solid densities. The resulting ion temperatures, 4–6 keV, should according to predictions allow the thermonuclear reaction



to proceed, with a very small contribution from the D–D reaction. Using such targets a group from the Lawrence Livermore Laboratory (LLL) reported at the 11th European Conference on Laser Interaction with Matter (Oxford, September 1977) that neutron yields in excess of 7×10^8 per laser shot had been achieved using the Argus neodymium glass laser facility.

Although the performance of such targets may not be very representative of ablatively driven systems they are of particular value in allowing detailed computer predictions of laser plasma interactions and magnetohydrodynamics to be checked experimentally. One question of fundamental importance which needs to be answered is whether these relatively high yields of neutrons

and α -particles are truly thermonuclear in origin or whether they are produced by fusion reactions between deuterium and tritium ions which have been accelerated to high energies within the plasma. Thermonuclear-produced particles are expected to be produced centrally in the compressed plasma whose ions have a maxwellian energy distribution at a temperature T_i (keV) giving rise to a gaussian-like energy distribution of neutrons and α -particles with a width (fwhm) of $\Delta E = 177\sqrt{T_i}$ (keV). Fusion products which are non-thermal in origin will have a different width ΔE depending on the details of the accelerating mechanisms. Early work by Rose *et al.* (*Nature* **181**, 1630; 1958) demonstrated the existence of non-thermonuclear fusion neutrons produced by the magnetic confinement device Zeta. McCall *et al.* (*Phys. Rev. Lett.* **30**, 1116; 1973) have postulated a detailed mechanism where the production of fast electrons by the resonance absorption of the laser light results in an electrostatic field on the outside of the compressed zone of the deuterium-tritium mixture. For nanosecond laser pulses a low density low temperature corona will be generated around this compressed zone before significant fast electron production occurs. As electrostatic acceleration occurs near the critical surface, where the plasma and laser light frequencies are equal, fast ions pass through the corona generating neutrons and alphas. For picosecond laser pulses, however, no low density plasma region is produced and the fast ions produce no particles. The non-thermonuclear particle production should therefore depend sensitively on the pulse characteristics and the comparisons with experiments made by McCall and others at Los Alamos would indeed seem to confirm this. The beam properties of the ions will now produce energy distributions of neutrons and alphas which differ significantly from the gaussian distribution associated with thermonuclear reactions. For example an equimolar monoenergetic D–T ion beam of energy E_i (keV) uniformly converging on a cold D–T target would result in a square distribution of neutrons and α -particles with a width $\Delta E = 150\sqrt{E_i}$ (keV).

Over the past 2 years a series of elegant experiments at LLL has confirmed the thermonuclear origin of the neutrons and α -particles produced in exploding pusher targets. The initial set of experiments (*Phys. Rev. Lett.* **35**, 1083; 1975) used a time-of-flight technique to obtain the energy distribution of the 3.52 MeV α -particles. By measuring the transit time $t = d/v$ over a flight path of length d of an α -particle of velocity v , the α -particles' energy dispersion $\Delta E = 2E(v/d)\Delta t$ is

obtained for a temporal resolution Δt . The experiment, designed so that α -particles in the range 2.6 to 3.9 MeV were magnetically deflected and detected by a scintillator, measured a dispersion consistent with a maxwellian temperature of 2.2 keV for the first laser shot and 3.0 keV for the second. This should be compared with maxwellian values in the range 1.6–3.0 keV calculated by numerically modelling the imploding microshells with a two dimensional Lagrangian hydrodynamics code. However such experiments are by no means conclusive since they assume that the α -particles are produced inside the glass microshells and consequently rely on a 0.2 MeV computed energy correction for the energy loss of the α as it passes through the compressed gas and the glass walls of the microsphere.

Consequently the Livermore group evolved an entirely novel approach to resolve the spatial distribution of the α -particles as they are born in the thermonuclear process by directly imaging the thermonuclear burn region within the target with a resolution of about 10 μm . (*Phys. Rev. Lett.* **39**, 22; 1977). The technique, which is essentially a holographic one, was originally proposed for stellar X-ray imaging and has a number of applications in nuclear medicine (*J. Nuclear Med.* **13**, 382; 1972). It proceeds in two stages. In the first the radiation source to be imaged casts a shadowgraph (a coded image) through a Fresnel zone plate (a coded aperture) onto a suitable detector, in this case a thick cellulose nitrate film. The image reconstruction is achieved at the second stage when a negative of the shadowgraph is illuminated with coherent light to produce a Fresnel diffraction pattern which is a reconstruction of the original α -source distribution but inverted and magnified. With associated neutron yield measurements this technique showed that more than 97% of the α -particles originated in a central region, ovoid in shape, which had a diameter about one-third that of the original target.

Final confirmation of the thermonuclear origins of the neutrons and, by inference, of the α -particles has now been obtained at Livermore (*Appl. Phys. Lett.* **31**, 645; 1977) using high resolution time-of-flight spectrometry on the neutron yield. Since there is negligible interaction of the emerging neutrons with the glass microsphere and its contents, measurement of the neutron energy distribution is a particularly useful diagnostic. Using 44.5-m flight paths with fast plastic scintillation detection, a mean measured neutron energy of 14.00 ± 0.10 MeV from five experiments compares well with the expected value of 14.05 MeV. Interpreted in terms of maxwellian ion temperatures the measured energy dis-

tributions correspond to a range of temperatures from 5.2 to 6.3 keV, in excellent agreement with the predicted thermonuclear value of 5 keV.

The prevalence of chaos

from N. MacDonald

It has been known since the time of Poincaré that quite simple conservative dynamical systems, such as two coupled linear oscillators, can exhibit highly complicated behaviour. As well as periodic trajectories, which are closed curves in phase space, there are also trajectories which, loosely speaking, fill up volumes in phase space. The rather special nature of conservative systems limits the number of physical contexts in which this behaviour can be manifested. (For some interesting examples see Henon & Heiles *Astr. J.* **69**, 73; 1964; Dunnett, Laing & Taylor *J. math. Phys.* **9**, 1819; 1968.) Analogous behaviour has been observed in recent years in a number of nonlinear models of physical systems which are open or dissipative, rather than conservative. This is leading to reappraisal of our understanding of these systems.

The best known example is that put forward by E. N. Lorenz (*J. atmos. Sci.* **20**, 130; 1963). He starts from the partial differential equations for flow in a layer of liquid of uniform depth with a constant temperature difference maintained between the upper and lower surfaces. These have a steady state solution with zero flow. As the temperature difference is raised this solution becomes unstable and convective flow solutions replace it. Lorenz examines one particular mode of convective flow, and uses a truncated Fourier expansion to reduce the problem to one involving three coupled ordinary differential equations. The variables are the intensity of convective flow, the temperature difference between ascending and descending currents, and a quantity measuring the departure from linearity of the temperature gradient across the fluid layer. Lorenz finds a second instability, beyond which the simple smooth convective flow is replaced by an extremely complex and unpredictable flow. The mathematics of this situation have become popular in the past few years, and progress in understanding the nature of the complex flow is reviewed in the proceedings of a seminar

N. MacDonald is in the Department of Natural Philosophy in the University of Glasgow.

The blind fishes of Persia

THE note published in *News and Views* (**264**, 113; 1976) on the blind cave fishes of Persia mentioned the discovery of the two cavernicolous forms *Noemachilus smithi* (Cobitidae) and *Iranocypris typhlops* (Cyprinidae) in the same well-like outlet. This is of great interest for the zoogeographical and etho-ecological interpretation of the possible mechanisms which led epigeic fishes to colonise subterranean waters and evolve underground into isolated homogeneous populations composed entirely of eyeless and depigmented individuals. These phenomena of regressive evolution have been well studied on several populations inhabiting caves (Thinès *L'évolution régressive des Poissons cavernicoles et abyssaux*, Masson, Paris, 1969), but little, if any, specific attention has been given to the case of cavernicolous forms found in wells and artificial outlets. It is therefore fortunate that Mr Smith was able to discover this new *Noemachilus* form in association with the previously discovered *Iranocypris* (Brunn & Kaiser *Dan. Scient. Invest. Iran*, Munksgaard, Copenhagen, 1943). However, the coexistence of two distinct regressed species in the same well has been observed before. In 1962, Mees found two different blind and depigmented Teleosts, *Milyeringa veritas* (Gobiidae) and *Anommatophasma candidum* (Synbranchidae) in two wells depending on the same subterranean system and located about 10 miles apart, near Yardee Creek Station, North West Cape, Australia (Mees *J. R. Soc. West. Austr.* **45**, (I), 24; 1962).

The coexistence of distinct cavernicolous forms in a same biotope other than a cave, raises the question of the exact nature and extent of their habitat. Underground waters can be traced with sufficient precision in many karstic systems (the system of sinks, underground caverns and streams found in limestone areas), whereas water sheets feeding wells or pumping sites do not lend themselves to similar descriptions. In addition, such phreatic water tables cannot be explored by speleologists, so that the only topographical data are those furnished by geological surveys. Thus, biospeleologists have had to limit themselves to inference as far as the

habitat is concerned. The only sound hypothesis about regressed fishes found in wells and other artificial water outlets is that these forms normally live in the phreatic sheets and are occasionally attracted to zones exposed to light by organic remnants of animal or human origin. The fact that fishes of the same species are sometimes observed in wells separated by a great distance is, in my opinion, the strongest argument in favour of this view. Mees reported the presence of the free-swimming *Milyeringa* in only two of the three wells he visited (*Milyeringa*, Kudmurra and Tandabiddi, the last halfway between and some 10 miles from the other two), whereas the eel-like *Anommatophasma* was found in all three sites. The cases of *Iranocypris typhlops* and of *Noemachilus smithi* seem similar to that of *Milyeringa veritas*. The ecological conditions in which these fishes live are moreover similar to those reported as early as 1904 by Goeldi for *Phreatobius cisternarum* (Goeldi *Rept. 6th Int. Zool. Congr. Bern*, 549; 1904), a very peculiar catfish belonging to the family Trichomycteridae which he found in a cistern of Marajo Island in the mouth of the Amazon and which was rediscovered recently (Delamare, personal communication). I have collected specimens of the blind Clariidae *Uegitglanis zammaranoi* in the canalizations of the Ischia-Baidoa pumping station (Somaliland) some 60 miles from the well of Uegit, where the first specimens were discovered in 1923 by Gianferrari (*Atti. Soc. Ital. Sc. Nat. Mus. Civ. Stor. Nat. Milano*, 62; 1923). To the present day, this distance is by far the greatest to have been reported for the geographical extension of a cavernicolous fish not dwelling in a cave (Thinès *Rev. Zool. Bot. Afric.* **57**, 117; 1958). The case of *Noemachilus smithi* can thus be considered as a typical instance of an already well-known phenomenon, whose meaning for biospeleological research is clear.

G. THINÈS

CPEC Animal Behaviour,
Louvain University
3041 Pellenberg
Belgium

on turbulence held at Berkeley in 1975 (Springer Lecture Notes in Mathematics No. 615; 1977.)

The conditions for this new instability are physically unattainable in this experiment. However, H. Haken has

pointed out (*Phys. Lett.* **53A**, 77; 1975) a remarkable parallel between the Lorenz model and a classical approximation to the dynamics of a single mode laser. In this case the three variables are the field strength and polar-

isation of the radiation, and the inversion of the atomic levels. The second instability can be attained if the laser cavity loss is high enough. Haken identifies this instability with the onset of undamped spiking of the laser output.

These models are purely deterministic, being set up in terms of ordinary differential equations. Yet the solutions behave in an apparently random way. The term 'chaos' has been appropriated to refer to this intriguing feature of nonlinear dynamical systems of three or more dimensions. A beginning has been made in the classification of different types of chaotic flow, and the determination of the conditions under which they occur. In a fascinating series of papers (Z. Nat. **31a**, 259, 1168, 1664; 1976; **32a**, 607; 1977) O. E. Rössler outlines some of the taxonomy of chaotic flows, and sketches a number of potential applications.

To visualise one possibility, think of a conical spiral seashell with a central pillar, and imagine a line winding out along the expanding spiral until it reaches the mouth of the shell, then tucking itself into the pillar and re-injecting into the narrowest coil. The equivalent behaviour is impossible for the trajectory of a dynamical system in two dimensions, because it would require that the trajectory cut itself at a point which is not a point of equilibrium. This kind of spiral trajectory could of course be periodic. The essential new thing about chaos is that for some sets of nonlinear equations all such periodic trajectories are unstable. The majority of trajectories then behave as if repelled, not only by the unstable focus at the point of the shell, but by each of the unstable periodic trajectories. Each time a trajectory re-enters to somewhere near the point it emerges in a different direction.

One may invent other varieties, for example two unstable foci with reinjection from the outside of each spiral to an inside coil of the other. H. Haken and A. Wunderlin (*Phys. Lett.* **62A**, 133; 1977) have given a simple geometric picture of the Lorenz-laser chaotic flow.

Rössler's extensions of the applications of chaotic flows have mainly been to schematic models of chemical reactions, involving oscillating and switching components. He has also discussed the relevance of this topic to the Rashevsky-Turing model of morphogenesis as controlled by chemical reactions and diffusion between cells. In a recent issue of this journal (*Nature* **271**, 189; 1978) Rössler and K. Wegmann propose the application of these methods to the Belousov-Zhabotinskii oscillating reaction (for a recent review see Tyson *The Belousov-Zhabotinskii Reaction* (Springer Lecture Notes in

Biomathematics No. 10; 1976)).

When this reaction is studied in a well-stirred solution, oscillations of the chemical concentrations, with well-defined period, can be detected by colour changes in an appropriate indicator. Also when the solution is in a shallow dish and unstirred, waves of colour change can be observed on the surface. A. Winfree suggested (personal communication to Rössler) that some of his observations on these waves might indicate symptoms of chaos. Rössler and Wegmann now report observations of the electrochemical potential in a well-stirred Belousov-Zhabotinskii reaction. This shows irregular spiking over times of the order of an hour, with about three spikes every 2 minutes. They suggest that this may be related to a particular type of chaos, which they illustrate by computed solutions of a very simple set of equations

$$dx/dt = -y - z$$

$$dy/dt = x + 0.55y$$

$$dz/dt = 2 - 4z - xz$$

Even this apparently innocent set of equations, in which there is only one nonlinear term, have rich possibilities. They can exhibit the 'seashell' type of chaos as well as the type which Rössler and Wegmann consider may be present in the Belousov-Zhabotinskii experiment. The authors point out that simultaneous measurements of two or more quantities can give a much more critical test of their hypothesis, and comment briefly that they have some promising results, plotting the electrochemical potential against the potential of an electrode sensitive to bromide ions.

Other authors (Schmitz, Graziani & Hudson, *J. Chem. Phys.* **67**, 3040; 1977) have reported data on the well-stirred Belousov-Zhabotinskii reaction consistent with a different type of chaotic flow while L. F. Olsen and R. Degn (*Nature* **267**, 177, 1977) have presented evidence for chaotic flow in the horse-radish peroxidase reaction. It is thus becoming apparent that even without the manifest complications brought in by considering diffusion as well as reactions, these systems reveal a rich variety of phenomena. This will undoubtedly stimulate much further work, both theoretical and experimental, on chemical oscillations and on their possible bearing on biological problems.

Natural fission reactors—Oklo style

from S. A. Durrani

A Technical Committee Meeting (Experts Group) on Natural Fission Reactors was held at Paris on 19–21 December 1977. It was organised by the International Atomic Energy Agency (IAEA) in collaboration with the French Atomic Energy Commission (CEA).

AFTER 5 years of investigation into the Oklo phenomenon, it is no longer a question of why it happened but why it should not have happened all over the world. Several researchers have been hunting for clues at likely places but as yet no other natural fission reactors have turned up.

The Oklo phenomenon (see *News and Views* **256**, 264; 1975), was first discovered by French scientists in 1972 when they established that a self-sustained chain reaction had occurred in a uranium mine, some 1,800 Myr ago, at Oklo in the Republic of Gabon in west equatorial Africa. The higher natural abundance of ^{235}U , namely over 3%, prevailing at that time (more than two half-lives ago for the isotope), the moderation of the fission neutrons by soil water (making it more likely that they would initiate new fission reactions), and effective absence of neutron-absorbing 'poisons', constituted favourable conditions for the occurrence of a self-propagating fission reaction. Many of the fission products, in proportions exactly the same as those obtained in today's man-made reactors, still lie there as evidence of the long-ago activity. That in itself was an important discovery, with far-reaching implications for the storage of nuclear waste.

The Paris meeting of the Technical Committee on Natural Fission Reactors (Experts Group), convened jointly by the International Atomic Energy Agency (IAEA) and the French Atomic Energy Commission (CEA), was the first of periodic conferences on the subject recommended to be held by a permanent international working group on Natural Fission Reactors (NFRs) set up under the auspices of the IAEA following the 1975 conference at Libreville in Gabon. Over 40 papers were presented on topics ranging from the geology and petrology of the Oklo reactor materials to isotopic chemistry, radiation damage studies of samples from a reaction zone and its surroundings, reactor physics

S. A. Durrani is a Senior Lecturer in the Department of Physics, University of Birmingham, UK.

models of the NFRs, and the role of organic materials in concentrating the initial uranium ore deposits.

The opening address was delivered by R. Naudet (CEN, Saclay), the scientific secretary of the meeting and since 1972 the director of the 'Franceville Project' established by the French CEA to promote and coordinate the international investigation of the NFR phenomenon. He reported on the 5 years of exploration of the Oklo site and of the international ramifications of the project. In the papers that followed, the emphasis was on natural fission reactors in the plural—not only at Oklo but elsewhere in the world. The development of the phenomenon was traced (or imagined) in both time and space. Investigations have shown that there was not just one reactor at Oklo: at least six 'reaction zones' over a relatively small area (a few thousand square metres) have been identified. A question exercising not a few minds was whether one of these was a 'primary' reactor and the others either secondaries or its lineal descendants. The reactors, like reproducing organisms, seem to have been 'propagating' themselves. Naudet, in true detective story style, termed this the 'Oklo mystery'. The age of the phenomenon has now been put back from ~1,800 Myr to nearer 2,000 Myr; the duration of the entire drama at Oklo, from the first spark to the last smolderings of the atomic fuel has also been extended from ~500,000 years to ~2 (or perhaps several) Myr.

A systematic search for other natural fission reactors is now being conducted elsewhere in the world, as described by K. E. Apt (Los Alamos Scientific Laboratory). Apt and his colleagues have analysed high-grade pre-Cambrian uranium ores from Australia, Canada, South America, Africa and the USA, for tell-tale isotopic variations in uranium and its fission products. No isotopic anomalies, indicating criticality, have so far been discovered in the samples from any of those sources. But, as was evident during discussion, everyone now expects another natural fission reactor to be discovered in some part of the world sooner or later. The hunt is on.

Amongst the papers on physico-chemical aspects of the phenomenon were those on estimation of the reactor temperature (~280–400 °C) by P. Holliger and colleagues (CEN, Saclay) and R. Vidale (Los Alamos Scientific Laboratory); gauging the adverse effect of intense radiation on the thermoluminescence properties of Oklo materials inside the reaction zone by S. A. Durrani and colleagues (University of Birmingham); problems of criticality and reactor control in Oklo and other NFRs by R. Naudet; and geochemical study of insoluble organic

material (kerogene) of the Oklo uraniferous ore by M. Vandenbroucke and colleagues (Institut Français du Pétrole).

A final round table discussion focused on the relevance of the Oklo experience to present-day problems of nuclear waste storage. G. A. Cowan (Los Alamos Scientific Laboratory) reported evidence for the remarkable stability of uraninite (UO₂) grains at the Oklo reactor site. He estimated fractional loss rates of 10⁻¹⁰ yr⁻¹ for elements such as uranium and plutonium. Even for the most volatile

elements, the loss rates during the operating period of the reactor were unlikely to have exceeded ~10⁻⁶ yr⁻¹. D. G. Brookins (University of New Mexico) demonstrated that many of the fission products (the rare earths, for example) had hardly migrated from the host pitchblende and its associated gangue over the past ~2 × 10⁹ yr. These observations greatly strengthen the case for the storage of long-lived nuclear waste products (such as plutonium) in tectonically stable geological repositories, possibly after incorporating them in synthetic uraninite matrices. □

Telescopes of the future

from Edward Kibblewhite and John Whelan

The ESO Conference on Optical Telescopes of the Future was held at CERN, Geneva on 12–15 December, 1977.

CONTINUOUS improvement of detectors has enabled astronomers to use the light collected by telescopes more efficiently and to observe fainter objects in shorter times without building bigger telescopes. L. Woltjer (ESO, Geneva), in opening the conference, pointed out that the 'collecting power' of telescopes (area of mirror times quantum efficiency of detector) has doubled every 7 years since 1910. Since the next generation of detectors will approach quantum efficiencies of unity in the near future, the only way to improve collecting power is to build telescopes of bigger effective mirror area.

However, I. King (ESO), R. Ekers (Gröningen) and M. Disney (University College, Cardiff) arguing from different standpoints, showed that, when observations are background-noise limited, the important scaling factor for signal-to-noise detection efficiency is diameter of telescope (D)/seeing disk size (S). Only high-resolution spectroscopy, which is photon noise limited, scales at D^2 . Some parts of the conference were concerned with increasing D and others with decreasing S . The cost of a telescope system (including dome, computer and so on) using current design scales as D^γ where γ was said to be between two and three; to build large telescopes therefore requires new designs to keep the cost down.

Two telescopes, which are to come into operation shortly, showed that low mirror weight can give great savings in

cost. J. Ring (Imperial College, London) described the new SRC 3.84-m infrared telescope being built in Hawaii. This uses a very thin mirror and a light mounting and cost £2.5 million, one-fifth to one-half the cost of a conventional telescope of the same aperture. P. Strittmatter and colleagues described the SAO/University of Arizona multiple-mirror telescope (MMT). This consists of a rigid frame holding six 1.8-m telescopes which converge the light to a common focus. Each telescope can be moved individually under computer control, and is aligned by looking at a bright star and analysing the image at the focus by computer. With all telescopes aligned, a 0.8-m telescope in the middle of the array, which also acts as a guider, produces a set of parallel laser beams which maintain the alignment of the six telescopes with respect to each other. The MMT will produce $\frac{1}{2}$ arc s images over a 1 arc min field of view, though it is primarily to be used for on-axis work (about 80% of the observations at a typical observatory are on-axis). The structure is alt-az mounted and the whole building rotates. The MMT, which contains many important innovations in telescope design, is said to cost \$8 million and has an equivalent aperture of 4.5 m.

Kitt Peak presented four designs for the 'next generation telescope' (NGT), each providing a 25-m equivalent aperture. A rough cost of \$10⁸ was mentioned—the cost of the Very Large Array (VLA). One design, the 'Shoe', involved a rotatable sector of a large spherical mirror, like part of the Arecibo radiotelescope, the spherical aberration being corrected by auxiliary optics. Other proposals were an enormous steerable dish made out of a mosaic of mirrors, an MMT design

Edward Kibblewhite and John Whelan are at the Institute of Astronomy, Cambridge.

with six 10-m telescopes, and the 'singles array', in which six 10-m telescopes would feed a central coude focus (see picture). The engineering problems of building a steerable telescope of 25-m diameter were felt to be serious; furthermore the large 'plate scale' (mm per arc s) makes the telescope incompatible with modern detectors. Simultaneous guiding of several smaller telescopes does not seem to be a problem; a suggestion to use accurate gyroscopes to achieve this was made.

Various siderostat telescopes were proposed which have a fixed telescope fed by a larger flat. A. Meinel (Optical Sciences Center, Arizona University) described one with a 45° flat and R. Bingham (Royal Greenwich Observatory) a neat alt-alt design. W. Richter (ESO) proposed a novel telescope which used a giant windmill, both to protect the undomed telescope from wind vibration and also to generate 300 kW of power! A continuing theme throughout the discussion, and emphasised particularly by E. Becklin (NASA, Hawaii) was the need to provide for adequate infrared capability.

At present, observing time on large telescopes is heavily oversubscribed. Disney has, for some years now, proposed incoherent arrays of telescopes in which information is collected by detectors and combined afterwards. He suggested a group of telescopes all identically instrumented which could either all point at one object or act independently. The availability of essentially noise-free detectors now makes this scheme very attractive from an astronomical point of view.

The importance of obtaining greater efficiency by improving the effective seeing disk size was not overlooked during the conference. One uncertainty concerning large (≈ 10 -m) mirrors is whether temperature effects across the mirror would worsen the seeing. Meinel claimed this was already significant. For the proposed space telescope the 'seeing' disk will be about one-tenth that of a good ground-based site and it is this feature which allows it to go so faint. Interesting papers were given by F. Dyson (Princeton University) and J. Hardy (ITEK Corp., Lexington) on the use of active optics to compensate for seeing by deforming the telescope mirror itself or the wavefront near the focus. Hardy showed data from a 1.5-m telescope in which active optics had reduced the seeing disk by a factor of two or three. The technique is probably restricted to bright stars.

The remainder of the meeting was devoted to ground-based optical interferometric techniques for obtaining ex-

Rift Valley fever

from Arie J. Zuckerman and David I. H. Simpson

RIFT Valley fever or enzootic hepatitis is a severe viral infection which primarily affects sheep and cattle causing many deaths in pregnant and newborn animals. The disease occurs naturally only in Africa and by 1912 it was recognised in the Rift Valley in Kenya. The infection is mosquito-borne and the virus, a member of the Bunyamwera group of arbovirus was isolated and identified in 1931. It was then also recognised that many infections occur in man, particularly by direct contact with sick animals or carcasses. Laboratory infections have also been reported. In 1950-51 an epidemic of the disease occurred in South Africa during which it is estimated that 20,000 persons became infected and 100,000 sheep and cattle died from the disease. Another severe epidemic occurred in the central regions in Southern Africa in 1975 (Van Velden *et al.* *S. Afr. med. J.* **51**, 867; 1977).

In October and November 1977 an outbreak of Rift Valley fever occurred for the first time in Egypt in the Nile delta. The number of human cases has been estimated at 10,000-20,000 with 70-80 deaths. In the areas most affected in the provinces of Sharqia Qalyub and Giza as many as 70% of the population were infected (*WHO Wkly. Epidem. Rec.* **53**, 1; 1978). The disease in man usually follows a transient febrile course with severe headache and bodily pains. Complications include haemorrhage, encephalitis and involvement of the eye with macular degeneration and temporary or permanent blindness. Haemorrhage generally indicates a

poor prognosis and autopsy of cases revealed necrosis of the liver associated with haemorrhage, tubular necrosis of the kidneys, pulmonary congestion and haemorrhages in the stomach and colon.

Experimentally lambs are highly susceptible and may die from massive necrosis of the liver within 36 h of infection. Mice die with hepatitis within 3 days, and other laboratory rodents are readily infected. Monkeys develop a mild fever. The pathological lesions are principally those of massive hepatitis in lambs and focal lesions of the liver in older sheep. There is often damage to kidneys, and there may be haemorrhages in the intestines and elsewhere.

In the entomological investigations undertaken in the affected areas in the recent epidemic in Egypt virtually only *Culex pipiens* was found and indeed the virus was isolated from this species of mosquito. People using mosquito nets or otherwise avoiding exposure to mosquitoes at night were apparently not affected. As before, contact and slaughter of sick animals is considered to have caused at least some human cases. While it will be extremely difficult to establish the exact origin of the outbreak it may have resulted from the smuggling of camels (*WHO Wkly. Epid. Rec. op. cit.*). This epidemic illustrates once more that viral infections know no boundaries.

Arie Zuckerman and David Simpson are in the Department of Medical Microbiology at the London School of Hygiene and Tropical Medicine.

tremely high spatial resolution. Hanbury Brown (University of Sydney) gave a critical review of intensity versus Michelson interferometry and concluded that, although his intensity interferometer had achieved all its aims it was probably not competitive with a Michelson for faint objects. This is mainly because the bandwidth of a Michelson is so much greater than that of an intensity interferometer. A. Labeyrie (CERCA, St Vallier-de-Hiery) has made great progress in this area and showed pictures and results of his Michelson interferometer which uses two 25-cm telescopes separated by 20 m. He also described the construction of novel and cheap 1.5-m tele-

scopes to be used in a much more ambitious set-up. C. Townes (University of California, Berkeley) described a heterodyning interferometer for use in the infrared. Several papers discussed various aspects of speckle interferometry which now appears to have become a standard technique.

A. Boksenberg (University College, London) gave a comprehensive and well-organised review of trends in detector development and some astronomical stories were related by J. Greenstein (Caltech). We were also treated to L. Goldberg's (Kitt Peak) visit to China and a tour of the 400 GeV accelerator, which rounded off an extremely successful conference.

review article

Plasma waves in the magnetosphere

D. J. Southwood*

A large variety of plasma wave phenomena are seen in the Earth's magnetosphere. Attempts at the theoretical explanation have had some successes, including wave induced loss of radiation belt particles and the Kelvin-Helmholtz instability source of geomagnetic pulsations. But there are also areas where theory needs more development: for example, on the ion wave turbulence seen on auroral magnetic flux tubes, the role of anomalous resistivity and the origin of the terrestrial kilometric radio waves.

THE Earth's magnetosphere provides a vast variety of plasma wave phenomena over nine decades of frequency ranging from hydromagnetic waves at 1 mHz, whose field at ground level was first detected more than one hundred years ago¹, up to the radio waves near 1 MHz which make the Earth at times the brightest radio planet in the solar system².

The plasma in the magnetosphere is collision-free and 'anomalous' transport effects may be caused by plasma waves and turbulence. Some waves, by contrast, are a secondary effect, but they are still interesting for the information they give about the state of the plasma that supports them. In fact because space exploration has provided much knowledge of the variety of background conditions, the magnetosphere has been a good testing ground of collision-free plasma theory³. There have been notable successes but also some surprises. Here we review some recent examples from both classes.

The background plasma

The plasma distribution in space and energy is crucial in determining what waves occur in a plasma and it is appropriate to start by a discussion of this.

The basic distributions of plasma in the magnetosphere can be rationalised by recognising the dominant role of solar wind driven convection. Because the charged particle Larmor orbits about **B** are much smaller scale than any gross feature except near the neutral sheet in the tail and at the magnetosphere boundary, everything is ordered as illustrated in Fig. 1 by magnetic flux tube. To a first approximation, any two particles remain gyrating about and bouncing back and forth along the same tube as each other throughout the convective motion. In fact we can think of the flux tubes themselves as moving, carrying particles along with them.

The magnetosphere forms an obstacle to the solar wind plasma streaming from the sun and a bow shock forms. Between the shock and the magnetopause boundary is the turbulent magnetosheath. Much of the sheath plasma is deflected around the magnetosphere but direct entry occurs in the cusp region and remnants of this plasma are swept back into the high latitude tail as the plasma mantle. Convection over the polar caps is tailwards, so that in the tail the plasma flows inwards towards its centre plane from both sides. In the centre of the tail is a weak field region, the neutral sheet, with a strong current. Field lines are 'broken' here by the reconnection process and emerging from this region with a sunward flow is a hot plasma with temperature 1–2 keV and density of a few particles cm⁻³ called the plasma sheet. This hot plasma trapped in the Earth's

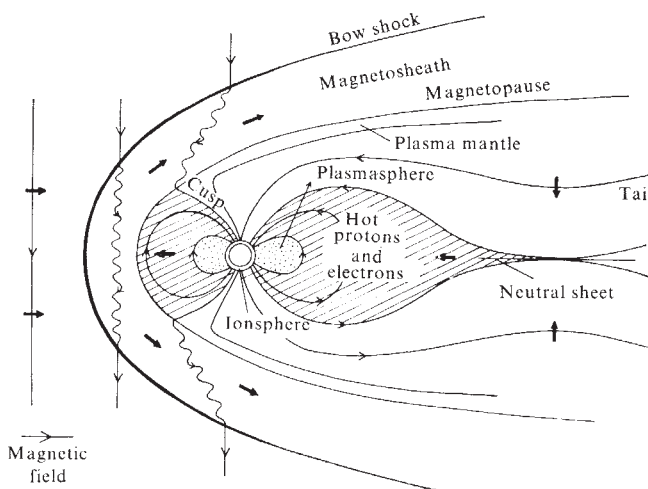
magnetic mirror field geometry flows inwards and drifts around the Earth to become the 'ring current' plasma, so called because of the depression in magnetic field it produces on the ground at the Earth. Throughout the ring current/plasma sheet the average proton energy exceeds the electron energy by about a factor of 2.

Within this region, in an ideal picture, is a region corotating with the Earth, the plasmasphere. This contains relatively cool plasma (temperature < 1 eV) with high density ($\sim 10^{-3}$ m⁻³) which is in equilibrium with the ionosphere. In a steady convection pattern ring current protons can penetrate the plasmasphere while hot electrons do not. This is because hot particles, as well as their convective 'flux tube' motion, have drift that is dependent on energy and charge.

As well as the ring current and plasmasphere populations the inner magnetosphere also contains the very energetic particles (energy up to MeVs in the inner zones) which constitute the Van Allen radiation belts. This energetic tail of the distribution plays little part in convection though their distribution is understood largely in terms of their injection and diffusion inwards by fluctuations in the convection system.

Several natural features of the convection system deserve

Fig. 1 An idealised view of the noon-midnight meridian magnetosphere. Broad arrows indicate the convective motion of flux tubes. The whole flux tube of plasma down to ionospheric heights moves together except in a region on the dayside magnetopause and in the tail neutral sheet where flux tubes are broken by a process called reconnection. The diagram is not to scale. The tail is very extended (~ 500 – $1,000 R_E$ (Earth radii)).



*Blackett Laboratory Imperial College of Science and Technology, Prince Consort Road, London SW7 UK

note at this point. Because flux tubes move as a whole, stresses are transmitted from the equatorial regions of field lines to the ionosphere and this is done by field aligned currents, 'Birkeland' currents, which may be carried by beams of particles. Birkeland currents have been detected near the outer boundary of the plasma sheet on high latitude field lines which map to the neutral sheet region, and it is also widely believed that the primary particles providing the discrete auroral displays (thin bright arcs) originate from this region. A second layer of Birkeland current oppositely directed to the high latitude set is found at lower latitudes and maps to the region occupied by the hot ring current/plasma sheet plasma. Plasma sheet electrons are regarded as the source of the steady 'drizzle' of electrons into the ionosphere that gives rise to the relatively structureless aurora, the diffuse aurora seen equatorward of the thin arcs.

The next point to note is that the convection system naturally provides both a heating of the hot plasma as it convects Earthward on the nightside and an increase in pressure anisotropy (pressure perpendicular to $\mathbf{B}$ > pressure parallel to $\mathbf{B}$). Finally a curious feature of energetic particle motion in the convection system is that the protons which best penetrate the plasmasphere are those with 10–20 keV energy, so that just inside the plasmasphere proton distribution functions occur with an absence of lower energy particles.

The convection can thus generate strong currents parallel to $\mathbf{B}$, particle beams, pressure anisotropies and particle distributions with a 'hole' at low energies (that is non-monotonic energy distributions). It also provides steep spatial gradients in particle distribution. Any of the preceding effects can provide a source of free energy which may drive instabilities or a steady spectrum of plasma waves. Often one envisages a quasi-steady situation where a relatively unchanging wave amplitude is sustained by convection or similar process providing energy at the rate wave energy is lost by some dissipative process such as bad ionospheric reflection.

Lastly note that convection is in fact 'gusty'. The most obvious manifestation is the substorm phenomenon when the auroral pattern breaks up, reconnection in the tail increases and convection and all its attendant phenomena intensify.

The above gives an outline of how the magnetospheric plasma distributions are set up, and as further background it is useful to note here some of the features of the characteristic frequencies that occur in plasma wave theory. The frequency at which protons gyrate about the magnetic field $\mathbf{B}$, the gyro- or cyclotron frequency, f_g^+ is proportional to the field B ($\approx 0.015 B$ Hz with B in gamma, $1\gamma = 10^{-9}T$) and in the equatorial plane ranges from a fraction of a Hz in the weak tail field to about 10 Hz in the inner magnetosphere. The corresponding electron cyclotron frequency, f_g^- ranges from a fraction of a kHz to a few tens of kHz in inner regions. The plasma frequency, f_p , the frequency at which electrons naturally oscillate about the ions in a cold plasma, is proportional to the square root of the electron density, n . In fact $f_p = 9\sqrt{n}$ Hz where n is the number of particles m^{-3} . f_p generally exceeds the electron cyclotron frequency but by typically less than an order of magnitude. In some regions off the equator, in particular a few Earth radii (R_e) above the auroral region ionosphere where much of the auroral kilometric radiation is believed to originate, the electron cyclotron frequency exceeds the plasma frequency.

This then sets the scene. The magnetosphere has hot and cold plasma populations, very largely collisionless and often far from thermodynamic equilibrium. Characteristic frequencies range from less than a Hz to hundreds of kHz (near the ionosphere). Waves are observed over the whole interval 1 mHz to about 1 MHz thus entirely containing the above range.

Waves around the electron cyclotron frequency

The word 'around' in the section title is to be interpreted loosely; the waves we look at in this section have frequencies from a few hundred Hz up to tens of kHz. We start by looking at frequencies below the electron cyclotron frequency. In a

cold plasma in this frequency range the wave that propagates is called the 'whistler' mode; the signal is in the acoustic band and recordings of signals which have come through the magnetosphere in this mode exhibit a characteristic whistling sound when played back through a tape recorder. Many whistler signals can be detected on the ground and some have an atmospheric source in lightning discharges.

It was recognised early on that whistler mode signals could strongly interact with energetic electrons¹. These fast electrons see the wave signal doppler shifted to their cyclotron frequency and a resonant interaction results. Even small amplitude waves can produce large orbit perturbations. Most significantly the pitch of the resonant particle's helical orbit along the magnetic field is altered and in the presence of a spectrum of waves a nett 'pitch angle diffusion' results. Particles with smaller pitch angles mirror further down the flux tube closer to the Earth and thus have an increased chance of being lost by collision in the atmosphere.

In the interaction electrons lose energy to the whistlers as their pitch angle is reduced and as a result a distribution with a preponderance of particles with large pitch angles is unstable to whistler waves. Such a pitch angle (and pressure) anisotropy is produced naturally by convection. Kennel and Petschek² pointed out that a steady state could be attained on a flux tube. Convection provided a source of particles while loss is by pitch angle diffusion. Waves gain energy because of the inherent instability but lose energy in bad ionospheric reflection. Development of a self-consistent theory still continues³.

In the magnetosphere it is interesting to note that two different kinds of whistler noise are commonly seen at high altitude. Outside the plasmasphere there is an emission called 'chorus'. Again named for its character, when replayed acoustically chorus resembles the dawn chorus of birds; the emission is sporadic, fluctuating on a time scale of seconds. Inside the plasmasphere the characteristic emission is a featureless 'plasmaspheric hiss' which effectively fills the high density region but is believed to be generated by cyclotron resonant interactions with >30 keV electrons near the outer edge⁴ of the plasmasphere. Detailed tests of the pitch angle diffusion theory have been made using typical wave fields and observed particle fluxes⁵ and it seems the basic mechanism is established. It is not established why chorus is structured in the way it is. The structure may well be allied to the nonlinear mechanism whereby signals from ground transmitters in the same frequency band generate sideband signals. Recently the Stanford group have shown that similar effects are achieved by very high harmonics of 60 Hz power line radiation entering the magnetosphere⁶ and have also shown that chorus occurrence may be related to power line radiation¹⁰.

An example of chorus and hiss occurrence is shown in Fig. 2 which shows data¹¹ from the Explorer 45 spacecraft. The spacecraft leaves the plasmasphere at 0424 UT (universal time). The hiss apparent in the channels below 1 kHz disappears at about this time and shortly afterwards chorus emissions appear in the 3.11 kHz channel for about an hour. It returns rather intermittently for about an hour before the re-entry of the spacecraft into the plasmasphere.

Also shown are emissions above the electron cyclotron frequency. In contrast to the whistler mode these signals have negligible magnetic signal and thus are electrostatic. First reported in 1970¹² these signals are specifically hot plasma phenomena which occur in bands between harmonics of the electron cyclotron frequency. They are regarded as responsible for pitch angle diffusion of the 1–10 keV electrons whose precipitation into the atmosphere provides the diffuse aurora. Simultaneous particle data from Explorer 45 (not shown) shows that the flux of these electrons is higher outside the plasmasphere than within it. The emissions are strongest in the band centred on $3f_g^-/2$. This is the characteristic band but near spacecraft apogee noise is present up to the $11f_g^-/2$ band. The theory of these waves and the associated precipitation is far short of the developed state of the whistler instability.

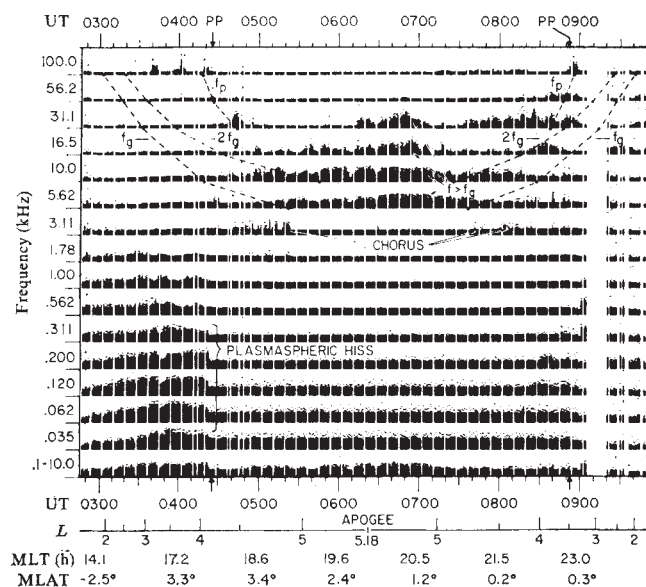


Fig. 2 Exp 45 (S3-A) spacecraft wave electric field data, orbit 214 day 23–23 January 1972. There are 15 narrow band channels and one broad band. Channel output is approximately proportional to the logarithm of the amplitude on a scale that varies from $2 \mu\text{V m}^{-1}$ to 20 mV m^{-1} . Vertical lines are 64 s averages while dots indicate peak amplitudes in the same period. f_g and f_p indicate electron gyrofrequency and plasma frequency. PP marks exit from the plasmasphere. L indicates the radial distance of the equatorial point of the magnetic field line through the spacecraft (in R_E). UT is essentially Greenwich mean time and MLT is local time. MLAT is latitude with respect to the geomagnetic equator. From Anderson and Maeda¹¹.

It is more complicated than the whistler instability discussed above for a variety of reasons. Whistler waves are relatively fast and pressure effects are unimportant. Here they are all important and complicated because wavelengths are comparable to the mean Larmor radius of the electron population. It is now agreed that waves are driven by pitch angle anisotropy of electrons and also that a weak cold background plasma is important¹³ and it was shown that the growth rate should peak just below the upper hybrid frequency, $(f_p^2 + f_g^2)^{1/2}$ (refs 14 and 15). The fact that observation shows that the waves just above f_g^- seem to be preferentially excited is resolved by noting that the group velocity is smaller for lower frequency waves and thus they can stay longer in the generation region.

The waves are intense with amplitudes of $1\text{--}10 \text{ mV m}^{-1}$ when compared with a background DC electric field associated with convection of a fraction of 1 mV m^{-1} in the magnetospheric equatorial plane and estimates¹⁶ suggest that their effect on particles they interact with are dramatic; 10 mV m^{-1} amplitudes are estimated adequate to provide pitch angle diffusion at a rate strong enough to fill the atmospheric loss cone in the time it takes a particle to move from one end of a flux tube to the other, the maximal 'strong diffusion' rate¹⁷. It is fair to say, however, that there is more theory and experiment to be done before we confidently understand the relationship between these waves and auroral precipitation.

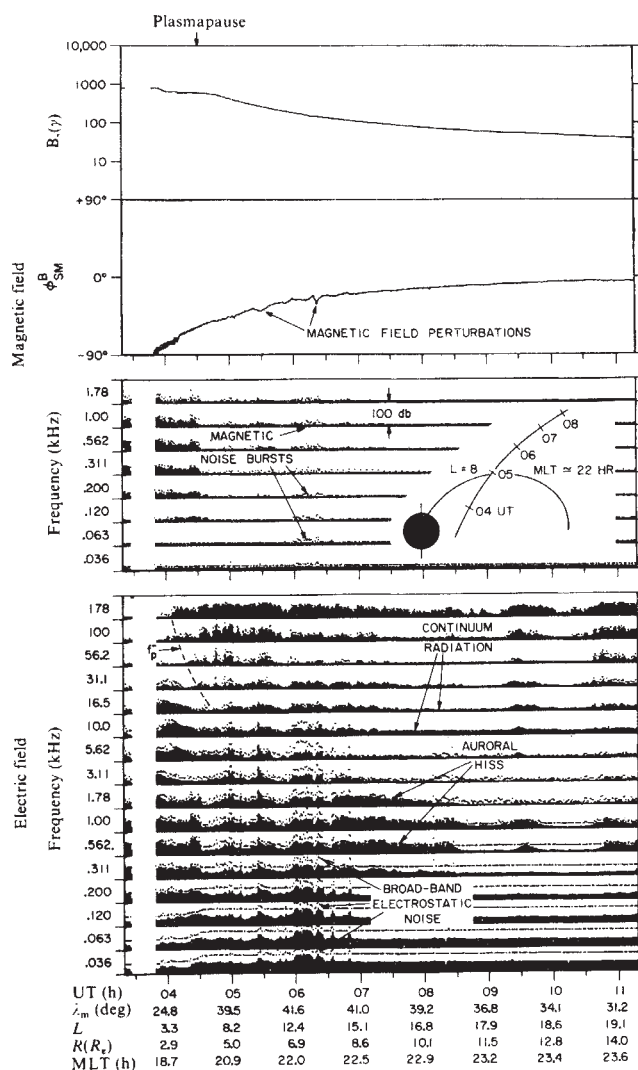
One feature of Fig. 2 deserves further comment: also shown is a noise band labelled by f_p . This plasma frequency noise is expected from cold plasma theory and it corresponds to the electron population oscillating with respect to the more massive ions. The oscillatory motion is along the magnetic field and as expected this electrostatic noise is polarised in the direction of $\mathbf{B}$. Another noise band is reported throughout the magnetosphere at the upper hybrid resonance frequency. Again predicted by cold plasma theory this band expands to take in the hot plasma $(n + \frac{1}{2})f_g^-$ waves in the hot electron region outside the plasmasphere. As expected the UHR noise is polarised perpendicular to $\mathbf{B}$ (ref. 18).

Ion waves

The right hand polarised whistler mode cuts off at the electron cyclotron frequency and in cold plasma theory there is a left hand polarised wave, the ion cyclotron wave, which cuts off at the proton cyclotron frequency. Energetic protons should be able to undergo gyroresonance with this wave mode in much the same way as electrons do with the whistler mode. Somewhat parallel theories were constructed⁵ and the decay of the ring current population in the recovery phase of a geomagnetic storm (a storm is a period when enhanced convection has shrunk the plasmasphere and increased the ring current population) is consistent with pitch angle diffusion loss induced by such waves¹⁹.

The waves are expected to occur at the high frequency end of the geomagnetic pulsation band (of which more later) and one type of emission in this band, called 'pearls', correlates with the recovery phase of storms²⁰. A second type of emission, intervals of pulsations with diminishing period (IPDP) is associated with substorms and has been identified simultaneously in space but without the decreasing period feature. This probably means

Fig. 3 Imp 6 spacecraft wave electric (lower panel) and magnetic field (middle panel) data in similar format to those in Fig. 2. The inset indicates the orbit in the late evening sector. The top panel shows background magnetic field strength and ϕ_{SM} is the azimuthal angle of the field in solar magnetospheric coordinates. For our purposes note that variations in ϕ are largely due to currents parallel to $\mathbf{B}$. λ_m is latitude with respect to the geomagnetic equator and R is radial distance from the Earth in R_E (Earth radii). From Gurnett and Frank²². Data are for 14 August 1971, day 226.



ground stations sample signals from a wider range of sources. Waves and particles have not been definitively correlated and some evidence that ion cyclotron waves are relatively infrequent²¹ may mean the instability is finely tuned. Theories agree that just inside the plasmapause is the place most conducive to instability and, as we have already mentioned, rather unusual proton distributions are produced there by steady convection. Ring current particles with energies below ~ 10 keV are absent and there is also a cut-off at high energy leaving particles with energies in a band ~ 10 –20 keV. If such distributions have developed the upper cut-off energy has a stabilising effect on wave growth while conversely the low energy hole in the distribution is destabilising. What is crucial is the contribution particles at upper and lower cut-offs make to the plasma pressure and this makes it feasible for such distributions to be stable to the waves.

The hole in the hot proton distribution predicted and observed at low energies would seem to be a good source of free energy for plasma waves but no correlation with waves has yet been reported. The distributions inevitably occur with the high density cold plasmasphere plasma as background and it is this cold background that inhibits the ion equivalent of the $(n + \frac{1}{2})f_g^-$ electrostatic waves discussed earlier.

Electrostatic ion waves above the ion cyclotron frequency are observed but outside the plasmasphere. Speculations on their existence had been made with the discovery of the electron waves¹² but discovery has been relatively recent²². These waves with frequencies from Hz up to kHz have peak power near 10–50 Hz and are found on flux tubes which map to auroral latitudes. The peak intensity is in the range of frequencies between the ion gyrofrequency and the lower hybrid frequency $(f_g + f_g^-)^{\frac{1}{2}}$ but extends up to f_g^- . These waves have amplitudes up to 10 mV m⁻¹.

At the upper end of the frequency range these waves are closely associated with an electromagnetic whistler mode emission called auroral hiss. Gurnett and Frank²³ report the hiss as extending outside the region of intense electrostatic signals which could indicate a common source, the electromagnetic signal being freer to propagate across the field. Downgoing hiss occurs in conjunction with intense 1–10 keV electron precipitation which in turn correlates with the discrete aurora structures and connections between this and the electrostatic waves remain an interesting source for speculation.

The signals occur over a wide range of local time and on the dayside the flux tubes containing the signals probably map to the polar cusp/magnetopause regions. The signals are found in the region where the higher latitude Birkeland currents flow and strong currents can be detected by the associated deflection of the Earth's magnetic field. Such deflections are seen within the electrostatic noise region and are associated with bursts of magnetic noise which seem to be propagating in the whistler mode but at a much lower frequency (~ 50 Hz) than the auroral hiss (kHz). Figure 3, taken from an outbound pass of the Imp 6 spacecraft on the nightside of Earth shows the electrostatic noise, auroral hiss and magnetic noise bursts²².

On the nightside the flux tubes on which the electrostatic noise occurs can reasonably be taken to map into the geomagnetic tail and it is appropriate next to mention that rather similar signals are seen in the tail regions^{23,24}. Figure 4 shows Imp 8 data²⁴ 30 Earth radii down the tail near the midnight meridian. The noise is present and seen in much the same frequency bands as the noise seen much closer to Earth. Once again it is found in the regions where Birkeland currents occur, the outer edge of the plasma sheet whose signature is the depression in magnetic field seen in upper panel. The electric vector is at a large angle to **B**. Whistler mode noise bursts are also seen within the electrostatic noise region and it seems that the waves in the two regions are closely connected. One significant difference is that the far weaker tail magnetic field ($\sim 5\gamma$) means the signals are between the local lower hybrid frequency and the electron gyrofrequency, the frequency range occupied at low

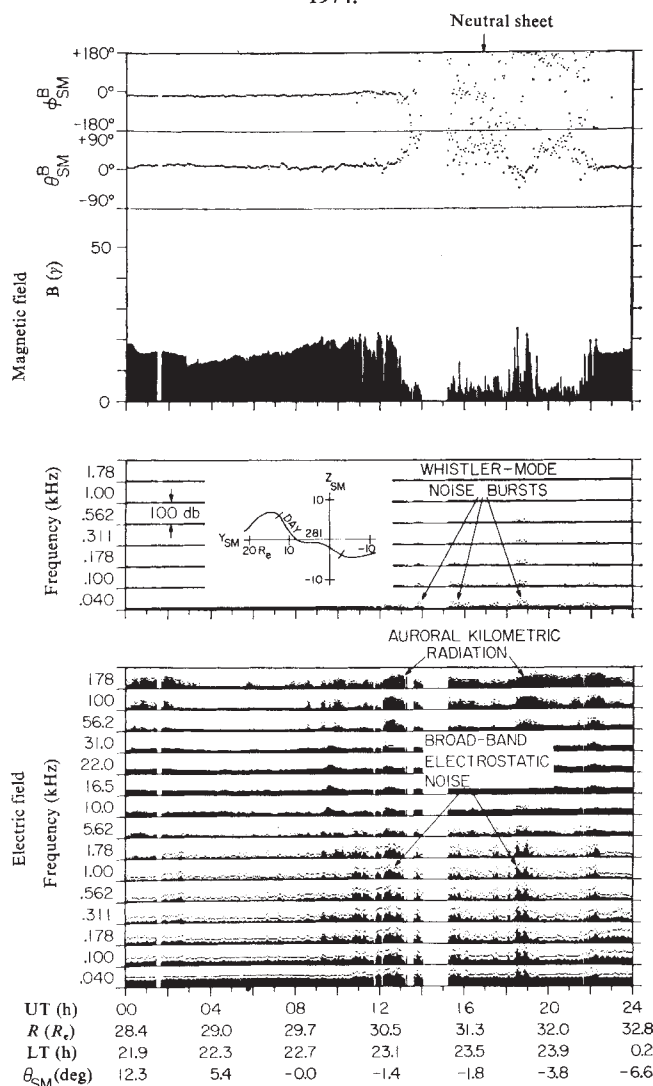
altitudes by the auroral hiss.

In the tail these signals correlate with high proton flow velocities and with auroral kilometric radiation intensity, as Fig. 4 shows. In turn the auroral kilometric radiation which we discuss later is strongly correlated with auroral activity.

Theory to describe these waves is still in the early stages of development. Several years before the observations were made Kindel and Kennel²⁵ predicted instability of ion cyclotron waves in the topside ionosphere ($\sim 1,000$ km altitude). These waves are driven by a current of cold electrons from the ionosphere. In fact the waves are seen at considerably higher altitude ($> 2,000$ km) and extend to very large radial distances.

Although a current driven source is tempting an alternative energy source is anisotropic pitch angle distributions of hot protons and there are theories predicting ion electrostatic waves in the outer magnetosphere^{26,27}. Ashour-Abdalla and Thorne²⁷ point out that these waves are strongly damped by the collisionless Landau damping process if a background cold electron population is present. If these particles are carrying a current such damping can be suppressed and the mere appearance of these waves at night and day in regions where large Birkeland currents are expected and seen argues strongly in favour of current driven instability.

Fig. 4 Imp 8 spacecraft electric and magnetic wave data in format similar to that of Figs 2 and 3 from 30 R_E down the tail. The inset indicates the orbit in a plane perpendicular to the Earth-Sun line. When B is depressed the spacecraft is in the plasmasheet. θ_{SM} and ϕ_{SM} are solar magnetospheric latitude and longitude. Noise is seen in conjunction with gradients in **B**. Nothing significant appears near the indicated neutral sheet crossing. From Gurnett *et al.*²⁴. Data from day 281, 8 October 1974.



Kilometric radiation

Also shown in Fig. 4 in the highest frequency channels is a signal labelled auroral kilometric radiation. These are part of Earth's radio spectrum^{2,28} which extends up to 1 MHz in frequency. As mentioned earlier these waves make the Earth at times a radio source as bright as Jupiter and the existence of these waves raises interesting theoretical problems. Gurnett² estimates that the radiated power may be as much as 10^9 W which can be placed in context by noting this is of order of 1–10% of the energy deposition rate associated with the aurora. Determination of source location by lunar occultation using the lunar orbiter RAE2 has revealed that the radiation largely originates from auroral zone field lines with most coming from altitudes between 1–2 Earth radii²⁹. There is a second source at high altitude in the cusp. The radiation correlates with auroral activity³⁰ and the flux changes by three orders of magnitude between quiet and disturbed times³¹.

These waves make a fascinating theoretical problem. There is no doubt they are in some way tapping energy available in auroral primary particles. Most theorists would agree on two points. The process is efficient ($\leq 10\%$ of auroral energy is converted into waves) and also the generation process is indirect in the sense I shall now describe. Radio waves propagate faster than light and thus it is argued cannot directly interact with a beam of auroral electrons. Rather a beam might be expected to generate electrostatic noise which then generates or is converted into the electromagnetic radio waves. Kaiser and Alexander³¹ show source frequencies are of order of the local upper hybrid and electron cyclotron frequencies and it is not unreasonable for a beam to generate electrostatic noise at such frequencies. Some theories³² then postulate conversion of the wave into a radio wave because of plasma inhomogeneity, conversion occurring when the wave frequency is near the local plasma frequency. An alternative theory³³ relies on large amplitude turbulence at the upper hybrid frequency to generate radio wave energy at twice the UHR frequency by nonlinear conversion. An ingenious idea of Palmadesso *et al.*³⁴ has a high frequency electrostatic wave generated by the beam interacting with a short wavelength low frequency ion wave to provide a high frequency long wavelength radio wave by wave-wave interaction.

All these theories have some disadvantages. On the basis of propagation cut-offs it is thought the waves are right hand polarised³⁵; the Palmadesso *et al.* theory predicts left hand polarisation. Very large amplitude electrostatic waves are required for nonlinear conversion while steep gradients are required for the most efficient conversion by inhomogeneity. There is so far little evidence of intense electrostatic noise in the appropriate frequency range though it can be argued that peculiarities of orbit coverage have left some of the expected source region relatively unexplored.

One alternative remains. It is possible for particles to see a radio wave doppler shifted to their cyclotron frequency and experience a direct gyroresonance interaction with the wave in this manner. This approach has been adopted by Melrose³⁶ who suggests the same mechanism may explain the similar radiation from Jupiter in the decametric band. Because the beam-wave interaction can be direct rather than a two-stage process it is appealing. The difficulty is that very extreme particle distributions appear to be needed if the electrons are to achieve a nett transfer of energy into the wave in the interaction.

The final story cannot then be said to be established. It will certainly involve the beams of electrons that are observed above the discrete aurora, the precise way the electrons in them are distributed in energy and pitch angle and probably also their distribution in space along with knowledge of the cold plasma distribution. It will also need to recognise the lower frequency hiss and ion waves that seem to originate in a similar vicinity. We discuss some further features of this region in the next section.

There is a second class of signal in the kilometric band as yet not definitely explained²⁸. This is the continuum radiation of much lower intensity and far more regular in character. Circumstantial evidence links this radiation with the plasma sheet electrons outside the plasmasphere and perhaps with the $3f_g/2$ electrostatic waves.

Anomalous resistivity

Anomalous resistivity remains a topic of great interest in the magnetosphere. There is circumstantial evidence of its occurrence, but much of the story waits to be developed. To understand what it is and its role, it is important to note that in a collisionless plasma one would expect the electrical conductivity along the magnetic field to be effectively infinite. The magnetospheric convection system requires currents to flow parallel to **B** as outlined earlier, though with more detailed structure expected on flux tubes that map to the neutral sheet region³⁷. Because of the high conductivity very small electric fields are required to give cold electrons a drift adequate to provide the requisite current. Once the current exceeds a threshold value, however, instability can set in. Energy from the drift motion is given up to plasma waves travelling at a similar speed to the current carrying particles. Naturally an increase in parallel electric field is then required to maintain the current. The effect of the instability is thus to increase the resistivity of the plasma.

In the presence of a turbulent wave spectrum the process is somewhat akin to the pitch angle diffusion process described earlier. Particles resonate with waves with phase speeds equal to their speed along **B** and the result of this strong interaction is that energy is fed to the wave while the particle diffuses in velocity. This process replaces the role of collisions in classical conductivity. If waves become strongly nonlinear the situation becomes far more complicated, wave-wave interactions may need to be considered and the radiation pressure of the waves can become important. Spatial structure is then produced by the exclusion of particles from regions where signals are most intense. Determining what stabilises a given turbulence spectrum is a taxing theoretical problem³⁸ but in the magnetosphere one test is to take an observed spectrum of waves and compute the particle scattering by the spectrum without considering the stabilisation problem. In fact given a wave spectrum regardless of the precise underlying instability one finds the effective collision frequency, ν^* , introduced by the turbulence is roughly given by

$$\nu^*/2\pi f_p \sim \epsilon_0 < E^2 > / 2p$$

That is, the ratio of effective collision frequency to plasma frequency is of order of the ratio of the wave energy and electron pressure. Such a formula was used by Fredericks *et al.*³⁹ in a detailed study of a double sheet of current found in the polar cusp region. Wave field strengths of 90 mV m^{-1} in a frequency band stretching from below 560 Hz up to several kHz were estimated to provide an anomalous resistivity 7.9 ohm m . Using the estimate of the current density deduced from spacecraft magnetometer measurements they estimate there existed a potential drop along the magnetic field of about 2 kV, enough to provide significant auroral acceleration and precipitation.

As the effective collision rate is proportional to the square of the turbulent wave amplitude and the amplitudes reported by the Iowa group near the Earth²² and in the tail²⁴ are at most a few tens of mV m^{-1} and a few mV m^{-1} respectively, resistivity due to the electrostatic turbulence emphasised in this paper is generally substantially lower. Gurnett and Frank²² suspect that in a region on appropriate flux tubes at about 1.8–4 Earth radii radial distance, as yet large unexplored, amplitudes would be higher. Some earlier reports suggest this is so⁴⁰.

Much interest centres on anomalous resistivity because of the uncertain role of parallel electric fields in acceleration of auroral

particles. In circumstances where a current of low energy particles is being disrupted, energetic particles can form a beam by running away in the parallel electric field³⁸. As mentioned earlier beams are seen in association with the low frequency noise. A recent observation⁴¹ in fact shows the noise in the presence of d.c. electric field structures suggesting a strong parallel field is present at the same time.

It is by no means a foregone conclusion that strong parallel electric fields are the prime auroral acceleration process. In fact at times the flux of energetic auroral primaries seems capable of carrying the observed Birkeland current. Causal relationships have still to be worked out in detail and another region where rapid particle acceleration is expected is the tail neutral sheet where, as briefly mentioned earlier, field lines undergo reconnection. Strong currents across the magnetic field are involved in this and one can picture the process as a conversion of magnetic energy into particle energy, but the proportions that go into thermal motions against directed motions (flows and beams) is not firmly established and probably variable. Observational information on flows in the tail is still a relatively recent development. Theories divide into quiet and turbulent current flow models. It does seem that macroscopically the flow is turbulent where the magnetic field is weak⁴² but there is a lack of any evidence of strong high frequency turbulence associated with current sheets. The only report⁴³ is of electrostatic waves of the $3 f_g/2$ type with low amplitude ($\sim 300 \mu\text{V m}^{-1}$ peak intensity). This could mean beam formation could occur in the neutral sheet region without much disruption.

Hydromagnetic waves

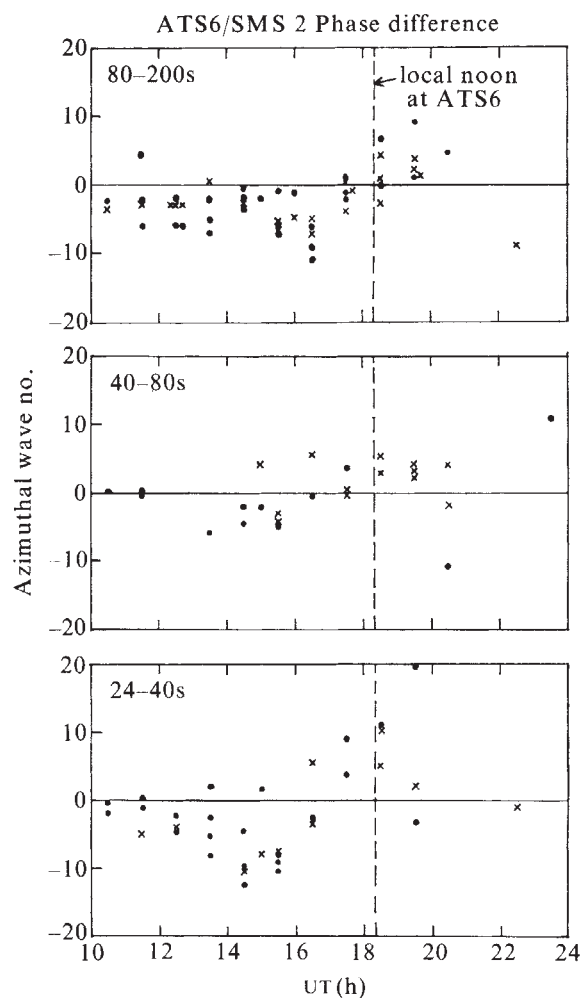
The phenomena discussed in this section are rather a contrast to those in the immediately preceding sections. Hydromagnetic waves are the lowest-frequency waves that are seen in the magnetosphere and a large number of them can be detected on the ground as geomagnetic pulsations, small fluctuations in the Earth's magnetic field. Their frequency is much below the proton cyclotron frequency and to a first approximation protons and electrons move together in the wave, the collisionless plasma behaving much like a fluid. One result of this is that waves can be described in hydrodynamic terms as well as electromagnetic. As with the background flow described earlier, the flux tube or field lines can be thought of as frozen into the wave flow and through the Maxwell stress tensor one can picture the forces which the magnetic field exerts on the plasma as a tension B^2/μ_0 along the field direction plus an isotropic pressure $B^2/2\mu_0$. A result of this approach is that one hydromagnetic wave mode can be recognised as directly analogous with waves on a stretched string. This mode, the Alfvén mode, has a magnetic perturbation perpendicular to $\mathbf{B}$, and thus bends the field without changing the magnetic pressure. These waves move along $\mathbf{B}$ at the Alfvén velocity $B/(\mu_0\rho)^{1/2}$, (ρ is mass density). The equatorial Alfvén velocity is of order 10^2 – 10^3 km s^{-1} in the magnetosphere while pulsation periods range from seconds up to ten minutes. One expects wavelengths comparable to the length of the field line (from ionosphere to ionosphere) for many of the signals seen and if this is so standing wave structures form along the field lines. As neither B nor ρ are uniform each field line (or more specifically field shell) has its own set of resonant frequencies. Such field line resonance signals eventually damp if only because the ionosphere does not perfectly reflect waves and the fact that the dayside ionosphere should reflect better than the nightside fits well with the observed morphology that continuous pulsations are more common by day and irregular damped pulsations more common at night⁴². However, other sinks of energy can be important including collisionless damping and also dayside and nightside energy sources do differ.

The best experimentally documented theory of dayside pulsation origin is that broadband hydromagnetic turbulence is set up near the magnetospheric boundary by the Kelvin–Helmholtz (wind-over-water) instability. The magnetosphere

then acts like a filter in that energy at a particular frequency is selectively amplified on the flux tubes that resonate at that frequency^{44,45}. The most recent evidence in favour of this source comes from magnetic field measurements made on three geosynchronous spacecraft, SMS1, SMS2, and ATS6 (ref. 46). This study showed that most signals show tailward-phase motion across the background field both in the morning and the afternoon, exactly as one would expect for waves driven by the solar wind. Some of the results are shown in Fig. 5. What is shown is the East–West angular wave number. There is a distinct tendency for the wave propagation sense to switch near local noon—best evident in the low frequency band.

Now the Hughes *et al.* study⁴⁶ is also important in showing the existence of a separate class of waves with short wavelengths across the background field. In Fig. 5 there is a gap in the data in the late afternoon (21–24 UT). Power was still present at this time of day but the signals showed little coherence between spacecraft, and thus phase difference measurement was not possible. One concludes that these signals have wavelengths shorter than the 4° longitude spacing of the geostationary spacecraft. Short wavelengths across the field are a characteristic of driftwaves, low frequency waves that occur in inhomogeneous plasmas, and these have received much attention in studies of laboratory plasma⁴⁷. Such waves are driven by extracting plasma internal energy where pressure

Fig. 5 The angular wave number perpendicular to the background magnetic field of hydromagnetic wave signals seen on two geostationary spacecraft ATS6 and SMS2. Positive wave number indicates eastward propagation and the change in sense near local midday indicates waves that are solar wind driven. The gap in the data in late local afternoon is significant and is discussed in the text. Dots and crosses distinguish polarisation differences not discussed here. From Hughes *et al.*⁴⁶.



gradients are present. There is little experimental evidence of the purely electrostatic modes that have received most attention in the laboratory context. Although measurement *in situ* of electric field is difficult at these frequencies such modes would cause particle flux oscillations and thus be detected. Electric field amplitudes of magnetic pulsation signals have been deduced in this way⁴⁸ directly confirming the standing wave structure described earlier. Back-to-back energetic proton detectors have detected wave variations on the scale of a Larmor radius⁴⁹. The particles involved in these observations are in the energetic tail of the distribution and as yet there has been no evidence that wavelengths across the field are as short as the thermal Larmor radius.

Hydromagnetic wave research can also be done on Earth relatively cheaply. Much of the field line resonance picture outlined above was developed in the light of data recorded on meridian chains of magnetometer stations on the ground. Curiously enough, measurements of East–West phase have been done only recently. Studies of waves using the British Institute of Geological Sciences magnetometer network at latitudes corresponding to inner magnetosphere flux tubes showed that their wavelengths were too long ($\sim 60^\circ$ of longitude) to be of drift wave origin while their propagation direction, preferentially eastward all day, was inconsistent with Kelvin–Helmholtz generation⁵⁰. In contrast recent high latitude measurements (Olson and Rostoker, personal communication) have shown consistency with the Kelvin–Helmholtz source like the geosynchronous spacecraft data.

Many pulsations may be driven by fluctuations in background convection and one class almost certainly is. These are the nightside ‘pi 2’ signals which at times can look like a text-book damped sine wave. An example of a train of such waves is shown in Fig. 6⁵¹. The upper trace was recorded near midnight in Newfoundland and shows a series of pi 2s while the lower trace was recorded in Lerwick, in the Shetlands, in the early morning. The more continuous Shetland pulsations show enhancements with the pi 2 bursts at midnight⁵¹. The midnight sector is where the most abrupt changes associated with substorms are seen. Pi 2 signals appear in close relation to the auroral intensifications, changes in convection and magnetospheric configuration. This was first pointed out ten years ago⁵² and one school now regards them as providing the best timing scheme for substorms⁵³. It seems reasonable to regard them as transients of the system associated with field lines coming to a new equilibrium with a change in convection conditions, like the twanging of telegraph wires in a gusty wind, and it would in fact be surprising if such did not occur. Their potential diagnostic use, however, is that they can appear outside the region where the most dramatic occurrences are. This is but one example of a plasma wave as a magnetospheric diagnostic. Other examples are legion and we could explore the subject much further.

Concluding remarks

This survey has not been exhaustive, but was designed to be illustrative of the type of knowledge we have. Instances of omissions are any mention of the magnetosheath, bow shock and near-Earth solar wind which closely relate to the magnetosphere and which each contain their own characteristic plasma waves known in a similar manner to those of the regions interior to the magnetosphere. We are also aware that magnetospheres other than Earth’s have their own wave phenomena, with Jovian radio emissions being a prime example. Direct measurements of plasma waves within the Jovian magnetosphere have yet to be made but no doubt will be. Moving to a cosmic scale both solar and pulsar radio phenomena have been suggested to have origins similar to terrestrial radiation. In the Earth’s magnetosphere we have the unparalleled advantage of direct sampling of the plasma in which the radiation originates, and of knowing the trapped plasma wave spectrum as well. Thus far these advantages have led to success-

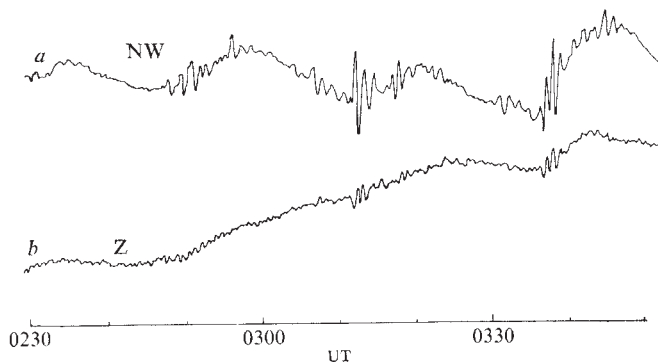


Fig. 6 Examples of nightside pulsation activity. Local time at Lerwick (b) is close to UT. St Anthony is $3\frac{1}{2}$ h behind and is near midnight. pi 2's at St Anthony (a) show as enhancements of low-level continuous activity at Lerwick. From Stuart and Booth⁵¹. Data from 21 August 1972.

ful developments of explanative collisionless plasma theory. One should also note that one often has more information about the plasma than in a laboratory plasma. Certainly the parameters a theorist wants to know (for example particle distribution in energy) are often very close to what is measured. There are, however, tantalising areas of exploration left. Often spacecraft have not flown with the right combination of instruments. International Magnetospheric Survey (IMS) activities on the ground and in space now underway should be productive. Worth a particular mention are the GEOS spacecraft of ESA and the NASA/ESA joint ISEE spacecraft projects. Both are well equipped to study waves and the latter is a multiple spacecraft project and as such the first of its kind. Happily, in the field of plasma waves alone there still remain many questions for them to answer.

1. Stewart, B. *Phil. Trans. R. Soc.* **423**, (1861).
2. Gurnett, D. A. *J. geophys. Res.* **79**, 4227 (1974).
3. Dungey, J. W. *Sci. Prog., Oxf.* **63**, 315 (1976).
4. Dungey, J. W. *Planet. Space Sci.* **11**, 591 (1963).
5. Kennel, C. F. & Petschek, H. J. *geophys. Res.* **71**, 1 (1966).
6. Eitcheto, J., Gendrin, R., Solomon, J. & Roux, A. J. *geophys. Res.* **78**, 8150 (1973).
7. Thorne, R. M., Church, S. R., Malloy, W. J. & Tsurutani, B. T. *J. geophys. Res.* **82**, 1585 (1977).
8. Lyons, L. R., Thorne, R. M. & Kennel, C. F. *J. geophys. Res.* **77**, 3455 (1972).
9. Helliwell, R. A., Katsufurakis, J. P., Bell, T. F. & Raghuram, B. *J. geophys. Res.* **80**, 4249 (1975).
10. Lurette, J. P., Park, C. G. & Helliwell, R. A. *Geophys. Res. Lett.* **4**, 275 (1977).
11. Anderson, R. R. & Maeda, K. *J. geophys. Res.* **82**, 135 (1977).
12. Kennel, C. F., Scarf, F. L., Fredericks, R. W., McGehee, T. G. & Coroniti, F. V. *J. geophys. Res.* **75**, 6136 (1970).
13. Young, T. S. T., Callen, J. D. & McCune, J. E. *J. geophys. Res.* **78**, 1082 (1973).
14. Karpman, V. I., Alekhin, Yu. K., Borisov, N. D. & Ryabova, N. A. *Plasma Physics* **17**, 937 (1975).
15. Ashour-Abdalla, M. & Kennel, C. F. in *Physics of the Hot Plasma in the Magnetosphere* (ed. Hultquist, P. & Stenflo, L.) (Plenum Press, New York, 1976).
16. Lyons, L. R. *J. geophys. Res.* **79**, 575 (1974).
17. Kennel, C. F. *Rev. Geophys.* **7**, 379 (1969).
18. Shaw, R. & Gurnett, D. A. *J. geophys. Res.* **80**, 4259 (1975).
19. Cornwall, J. M., Coroniti, F. V. & Thorne, R. M. *J. geophys. Res.* **75**, 4699 (1970).
20. Campbell, W. H. in *Physics of Geomagnetic Phenomena* (ed. Matsushita, S. & Campbell, W. H.), 821 (Academic, New York, 1968).
21. Kintner, P. M. & Gurnett, D. A. *J. geophys. Res.* **82**, 2314 (1977).
22. Gurnett, D. A. & Frank, L. A. *J. geophys. Res.* **82**, 1031 (1977).
23. Scarf, F. L., Frank, L. A., Ackerson, K. L. & Lepping, R. P. *Geophys. Res. Lett.* **1**, 189 (1974).
24. Gurnett, D. A., Frank, L. A. & Lepping, R. P. *J. geophys. Res.* **81**, 6059 (1976).
25. Kindel, J. M. & Kennel, C. F. *J. geophys. Res.* **76**, 3055 (1971).
26. Coroniti, F. V., Fredericks, R. W. & White, R. J. *geophys. Res.* **77**, 6243 (1972).
27. Ashour-Abdalla, M. & Thorne, R. M. *Geophys. Res. Lett.* **4**, 45 (1977).
28. Gurnett, D. A. *J. geophys. Res.* **80**, 2751 (1975).
29. Kaiser, M. L. & Alexander, J. A. *Geophys. Res. Lett.* **3**, 37 (1976).
30. Kurth, W. S., Baumback, M. M. & Gurnett, D. A. *J. geophys. Res.* **80**, 2764 (1975).
31. Kaiser, M. L. & Alexander, J. A. *J. geophys. Res.* **82**, 3273 (1977).
32. Jones, D. *Geophys. Res. Lett.* **4**, 121 (1977).
33. Barbosa, D. D. Thesis, Univ. California (1976).
34. Palmadesso, P., Coffey, T. P., Ossakow, S. L. & Papadopoulos, K. *J. geophys. Res.* **81**, 1762 (1976).
35. Green, J. L., Gurnett, D. A. & Shawhan, S. D. *J. geophys. Res.* **82**, 1825 (1977).
36. Melrose, D. B. *Astrophys. J.* **207**, 651 (1976).
37. Dungey, J. W. in *The Magnetosphere of Earth and Jupiter* (ed. Formisano, V.), 197 (D. Riedel, Dordrecht-Holland, 1975).
38. Papadopoulos, K. *Rev. Geophys. Space Phys.* **15**, 113 (1977).
39. Fredericks, R. W., Scarf, F. L. & Russell, C. T. *J. geophys. Res.* **78**, 2133 (1973).
40. Scarf, F. L., Fredericks, R. W., Russell, C. T., Kivelson, M. G., Neugebauer, M. & Chappell, C. R. *J. geophys. Res.* **78**, 2150 (1973).
41. Mozer, F. S. *et al. Phys. Rev. Lett.* **38**, 292 (1977).
42. Coroniti, F. V., Scarf, F. L., Frank, L. A. & Lepping, R. P. *Geophys. Res. Lett.* **4**, 219 (1977).

43. Newton, R. S., Southwood, D. J. & Hughes, W. J. *Planet Space Sci.* (in the press).
 44. Southwood, D. J. *Planet Space Sci.* 22, 483 (1974).
 45. Chen, L. & Hasegawa, A. *J. geophys. Res.* 79, 1024 (1974).
 46. Hughes, W. J., McPherron, R. L. & Barfield, J. N. *J. geophys. Res.* (in the press).
 47. Coppi, B. & Rewoldt, G. in *Advances in Plasma Physics*, 6 (ed. Simon, A. & Thompson, W.) 421 (Wiley, New York, 1975).

48. Kivelson, M. G. *J. atmos. Terr. Phys.* 38, 1115 (1976).
 49. Su, S.-Y., Konradi, A. & Fritz, T. A. *J. geophys. Res.* 82, 1859 (1977).
 50. Green, C. A. *Planet Space Sci.* 24, 79 (1976).
 51. Stuart, W. F. & Booth, D. C. *J. atmos. Terr. Phys.* 35, 2101 (1973).
 52. Rostoker, G. *J. geophys. Res.* 73, 4217 (1968).
 53. *IAGA Symposium SIII 3a, Timing substorm events* (abstracts), *EOS, Trans. Amer. Geophys. Un.* 58, 757 (1977).

articles

Limits to the expansion of Earth, Moon, Mars and Mercury and to changes in the gravitational constant

M. W. McElhinny, S. R. Taylor & D. J. Stevenson

Research School of Earth Sciences, Australian National University, Canberra, A.C.T., Australia

New estimates of the palaeoradius of the Earth for the past 400 Myr from palaeomagnetic data limit possible expansion to less than 0.8%, sufficient to exclude any current theory of Earth expansion. The lunar surface has remained static for 4,000 Myr with possible expansion limited to 0.06%, the martian surface suggests a small possible expansion of 0.6% while the surface of Mercury supports a small contraction. Observations of Mercury, together with reasonable assumptions about its internal structure, indicate that G decreases at a rate of less than $8 \times 10^{-12} \text{ yr}^{-1}$, in constant mass cosmologies, and $2.5 \times 10^{-11} \text{ yr}^{-1}$ in Dirac's multiplicative creation cosmology.

THE proposition that the Earth has expanded through geological time has received new impetus from those who claim that plate tectonics do not adequately account for many aspects of geology¹⁻³. It has been proposed that seafloor spreading exceeded subduction by 33% during the past 165 Myr, corresponding to an expansion of the Earth of over 10% since the Jurassic⁴. The possibility that the Earth or any of the planets have expanded has profound implications in cosmological theory because such expansion is predicted by models that suppose that the gravitational constant G has been decreasing with time⁵⁻⁸. It is not possible to calculate any possible changes in the Earth's radius from an examination of its surface features because these have been continually reshaped by a variety of geological processes. The only viable technique is that which uses the results of palaeomagnetic studies. In the past the number, reliability and lack of coverage of the Earth's surface of these measurements has made it difficult to make palaeoradius calculations with any accuracy. Successive analyses, however, have tended to eliminate all but the slowest ($< 2.5 \text{ mm yr}^{-1}$) expansion rates⁹. The situation has improved so much over the past decade that estimates can now, in favourable situations, be made with an accuracy of a few per cent. On the other hand, the results of the lunar exploration programme have shown that the surface of the Moon has remained relatively unchanged for nearly 4,000 Myr. Observations of the surface of Mercury and Mars by spacecraft enable constraints to be placed on the palaeoradius of these extraterrestrial bodies. Here we shall analyse palaeomagnetic data to provide constraints on the past changes in the Earth's radius and then look at the available data for the Moon, Mars and Mercury to limit possible changes that could have occurred in these bodies.

Palaeoradius of the Earth

Egyed^{10,11} first pointed out that palaeomagnetic data could be used to calculate the Earth's palaeoradius for different geological epochs. He showed¹⁰ that if two widely separated palaeomagnetic sampling sites were available for the same geological period and situated on the same palaeomeridian, the past latitude separation could be calculated from the observed palaeomagnetic inclination at each site. The ratio of the present angular separation to the past angular separation (determined by the palaeolatitude difference) was then the ratio of the past to present radius. Generally, of course, sampling sites are rarely on the same palaeomeridian and Egyed later suggested a more general triangulation method¹¹ that allows sampling sites to be arbitrarily located. Van Hilten¹² first used this method in a much modified form and concluded that a substantial increase in the Earth's radius had occurred since the Permian. A closer analysis¹³ of Van Hilten's method, however, has shown that his technique is incorrect and produces results that are at variance with his model. Van Andel and Hospers¹⁴ later proposed an improved triangulation technique and subsequently reviewed all estimates of the Earth's palaeoradius⁹ concluding that hypotheses involving large expansion rates^{15,16} could be rejected.

Ward^{17,18} has proposed a method which is appropriate to the spherical environment of the palaeomagnetic data. The analysis uses all data from a continental block in a single calculation. The dispersion of palaeomagnetic poles is investigated using Fisher's¹⁹ analysis of dispersion on a sphere for different values of Earth radii. That radius for which the dispersion of poles is a minimum is regarded as the best estimate of the palaeoradius for that data set. All results up to the end of 1976 and which conform with the minimum criteria for reliability set out by McElhinny²⁰ have been analysed in a new global review of palaeomagnetic data. The basic data set consists of the coordinates (λ_n, ϕ_n) of n rock units each having palaeodeclination and inclination (D_n, I_n). The Fisher¹⁹ mean position (λ_0, ϕ_0) of all rock units is calculated and the basic data are transformed to a new coordinate system with (λ_0, ϕ_0) as the pole to yield new values ($\lambda_m, \phi_m, D_m, I_m$) in this system. In this coordinate system, as the radius of the earth is changed (the continent keeping the same physical dimensions) so the latitude of the rock units (λ_m) will change to some new value (λ'_m) but the longitudes (ϕ_m) will be unchanged. For each value of radius new values λ'_m may be calculated and the resulting dispersion of the palaeomagnetic poles determined.

Table 1 lists calculations for several continents for those

Table 1 Estimates (R_a) of the Earth's palaeoradius as a fraction of its present radius

Block	Age	<i>N</i>	<i>D</i> _{max}	<i>K</i> _{max}	Data centre	Fixed centre	<i>R</i> _d	<i>R</i> _f	<i>R</i> _a
Africa	TR-J	18	70.7°	86.3	5.1°N, 9.1°E	5°N, 25°E	1.04	1.02	1.03
Africa	K	11	66.2°	35.6	13.2°N, 17.7°E	5°N, 25°E	(1.05)	(1.01)	(1.03)
Africa/Arabia	K	16	69.4°	29.3	13.2°N, 23.3°E	10°N, 30°E	0.99	0.99	0.99
Africa/S. America	P-C	10	64.8°	103.0	12.0°S, 1.5°E	10°S, 10°E	1.07	1.07	1.07
S. America	P-TR	10	40.9°	87.0	27.2°S, 61.7°W	15°S, 60°W	(1.21)	(1.17)	(1.19)
Africa/S. America	P-TR	13	55.0°	85.2	24.7°S, 7.5°E	10°S, 10°E	1.05	1.14	1.09
N. America	C	14	34.8°	*	45.6°N, 71.3°W	45°N, 100°W	*	*	—
N. America	P	14	37.0°	110.0	38.9°N, 100.7°W	45°N, 100°W	*	1.08	—
N. America	TR	28	38.1°	57.5	43.1°N, 89.0°W	45°N, 100°W	1.03	1.05	1.04
N. America/Greenland	J	10	45.3°	36.6	42.6°N, 88.1°W	45°N, 100°W	0.95	0.96	0.95
N. America	K	12	52.8°	65.0	44.5°N, 103.1°W	45°N, 100°W	0.98	0.98	0.98
Europe	D	24	40.1°	62.4	57.5°N, 40.1°E	55°N, 25°E	0.93	0.96	0.94
Europe	C	49	46.7°	57.5	54.6°N, 18.3°E	55°N, 25°E	0.99	0.99	0.99
Europe	P	63	41.1°	91.2	53.6°N, 29.7°E	55°N, 25°E	1.17	1.16	1.16
Europe/Asia	J	12	42.3°	33.8	46.3°N, 51.9°E	50°N, 80°E	1.00	0.95	0.98
Europe	K	14	60.8°	48.7	45.4°N, 49.5°E	55°N, 25°E	0.90	0.88	0.89
Asia (Siberia)	TR	36	32.6°	51.0	66.7°N, 100.8°E	60°N, 95°E	0.96	0.97	0.96
Asia	K	14	42.1°	42.7	58.9°N, 130.2°E	60°N, 95°E	1.03	1.03	1.03

R_a using palaeomagnetic data and Ward's^{17,18} method. R_d is the estimate using the average site location (data centre) as the pole for the coordinate system used in the calculations and R_f is the estimate using the approximate centre of the continental block as pole. N is the number of palaeomagnetic results used, $D_{\max}$ is the largest great circle distance (in degrees) between any two palaeomagnetic locations and $K_{\max}$ is the maximum value of Fisher's¹⁹ precision parameter for the palaeomagnetic poles. * No solution is possible.

Continental combinations use standard reconstructions^{22,23}. The values in brackets include data used in the larger set immediately following in each case. Calculations have only been listed where 10 or more locations are available.

geological periods where the palaeomagnetic results number 10 or more. One criticism of Ward's method is that the mean position of the rock units investigated is taken as the point at which strain due to curvature changes is zero. This might be far removed from the actual central point of the continent. The calculations have in each case been repeated using an approximate fixed central point in each continent. The differences between R_d (data centre) and R_f (fixed centre) calculations are negligible. So as to be unbiased, the mean, R_a , of these two values is taken as the best estimate in each case. In some instances Ward's method gives no solution because the data have been collected from too small a region of the continent. In many instances the change of the Fisher precision parameter K with radius is small and the solution for the ancient radius is not well determined. The best solutions in Table 1 are those in which there is the largest distance between the most widely separated rock units coupled with the largest number of observations. Even so the best of these determinations, probably from the Triassic–Jurassic of Africa, is only accurate to about 10% (ref. 21).

The data of Table 1 exhibit no systematic variation of radius with time from Devonian through to the present. The average radius over the past 400 Myr expressed as a function of the present radius is calculated as 1.01 ± 0.03 at the 95% confidence level taking the mean of the 14 independent estimates listed in Table 1. We can, however, improve the accuracy of the individual determinations by combining the data into larger sets, such as combining the data for Europe and Asia in post-Permian times or else reconstructing the continents into larger blocks that are known to have existed in the past. The first six entries of Table 2 involve a reconstruction of the Atlantic²² (Europe–North Amer-

ica, or Africa–South America) about which there is little dispute. Each of the six entries involves more than 25 rock units with a spread of at least 6,000 km between the most distant ones. The accuracy of each of these estimates of palaeoradius is probably about 5% at the 95% confidence level. The lower half of Table 2 gives calculations based on the Smith and Hallam²³ reconstruction of Gondwanaland and for the Triassic–Jurassic combining also with North America²². The accuracy associated with these calculations depends on one's view of the precision of the reconstruction used.

The first six entries of Table 2 provide the best overall estimates of the palaeoradius from all the palaeomagnetic data. Figures 1 and 2 illustrate the solutions for the presently largest single block (Eurasia) and for a reconstructed block (Africa–South America). Once again the results indicate no systematic change of radius with time since the Devonian. For the past 400 Myr the average palaeoradius may be calculated as 1.020 ± 0.028 (95% confidence limits). This value suggests a slight contraction of the Earth rather than an expansion. Any expansion is limited to less than 0.8% in 400 Myr. This corresponds to a maximum possible expansion rate of 0.13 mm yr^{-1} , a value sufficiently small not only to exclude the fast expansion of rates proposed by Carey^{2,15} and Hilgenberg¹⁶, but also to exclude the much slower rates proposed by Egyed²⁴ (1 mm yr^{-1}) or Wesson²⁵ (0.6 mm yr^{-1}). The data are marginally consistent with the contracting Earth models of Lyttleton⁵⁰, but the phase-change hypothesis in Lyttleton's theory is inconsistent with present knowledge of the equation of state of dense matter⁵¹, and the enormous latent heat required (equivalent to a temperature drop of 10,000K) was incorrectly omitted in his thermal evolution calculations.

Table 2 R_a of the Earth's palaeoradius using large combinations of palaeomagnetic data

Block	Age	<i>N</i>	<i>D</i> _{max}	<i>K</i> _{max}	Data centre	Fixed centre	<i>R</i> _d	<i>R</i> _f	<i>R</i> _a
N. America/Europe	D	30	67.9°	56.3	58.6°N, 23.6°E	50°N, 10°W	1.04	1.06	1.05
N. America/Europe	C	63	82.0°	63.1	54.5°N, 5.4°E	50°N, 10°W	0.98	0.99	0.99
N. America/Europe	P	77	77.4°	89.9	57.0°N, 12.9°E	50°N, 10°W	0.95	0.98	0.97
Eurasia	TR	58	69.5°	55.1	66.5°N, 67.4°E	55°N, 60°E	1.02	1.03	1.03
Eurasia	K	28	95.9°	46.0	59.1°N, 82.2°E	50°N, 80°E	1.06	1.07	1.06
Africa/S. America	P-J	35	77.9°	62.4	10.0°S, 6.1°E	10°S, 10°E	1.03	1.02	1.02
Gondwana	P-C	16	116.9°	100.6	23.3°S, 25.5°E	15°S, 40°E	1.07	1.07	1.07
Gondwana	P-TR	23	107.6°	43.7	25.6°S, 32.0°E	15°S, 40°E	1.03	1.03	1.03
Gondwana	TR-J	35	113.8°	74.6	13.0°S, 26.8°E	15°S, 40°E	0.99	0.98	0.98
Africa/S & N. America	TR-J	43	85.5°	61.8	13.7°N, 9.5°W	0°N, 0°E	1.07	1.07	1.07
Gondwana/N. America	TR-J	56	130.7°	61.9	4.2°N, 5.7°E	10°S, 30°E	1.03	1.02	1.03

Symbols as in Table 1. Reconstructions of Bullard *et al.*²² and Smith and Hallam²³ used.

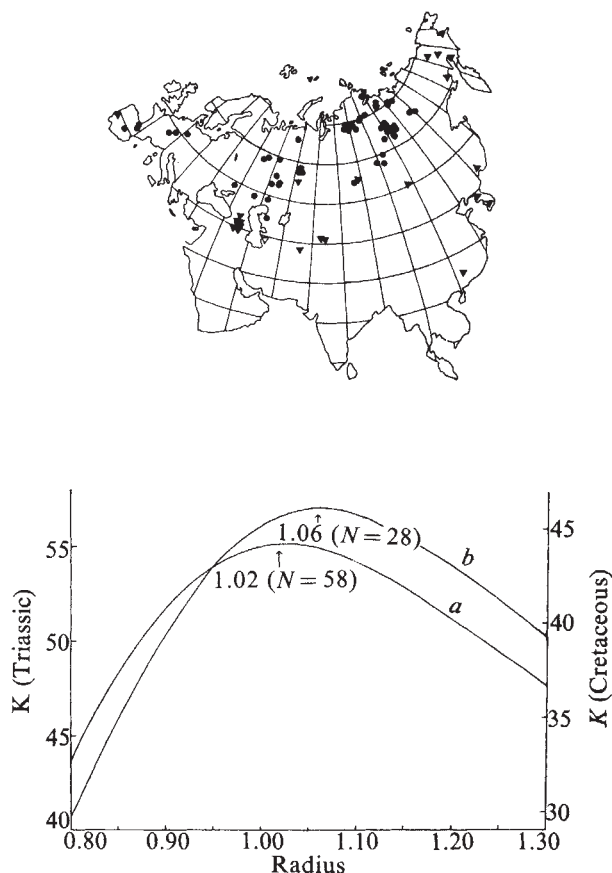


Fig. 1 Calculation of the palaeoradius for *a*, Triassic and *b*, Cretaceous palaeomagnetic data from Eurasia. The variation of the Fisher¹⁹ precision parameter *K* for the palaeomagnetic poles is shown as a function of radius (past/present) below the locality map (●, Triassic localities; ▲, Cretaceous localities).

Moon, Mars and Mercury

The surface of the Earth is not an ideal location in which to seek evidence of expansion. Although the geological record extends to 3,700 Myr, plate tectonic activity in the past 200 Myr is chiefly responsible for the presently observed surface structures. The inherent difficulties in interpreting the geological record are well known, and it is salutary to recall that the fleeting nature of the present landscape became an accepted fact less than 150 yr ago^{26,27} while the dynamic processes of recent seafloor spreading and continental movement became established only within the past 15 yr.

In contrast, the surface of the moon has remained apparently static for more than 3,000 Myr. Telescopic and orbital mapping, aided by manned lunar landings have established the essential features of the lunar surface^{28,29}. The terrain is divisible into two major areas, the maria and the highlands. The latter form an older heavily cratered surface. The origin of the large craters by meteoritic or asteroidal impact is now established. Evidence for the bombardment of the lunar surface by larger (~50 km diameter) objects is provided by the recognition of over 40 large (>220 km diameter) ringed basins, saturating the highland surface. Similar basins occur on Mercury³⁰ (for example, Caloris basin) and Mars³¹ (for example, Hellas, Argyre). Following the excavation of the basins, basaltic lavas flooded 17% of the lunar surface³². Since then, the only visible changes have been the production of younger impact craters (such as Tycho, Copernicus, Kepler) of which 15 are visible telescopically on the Earth-facing side.

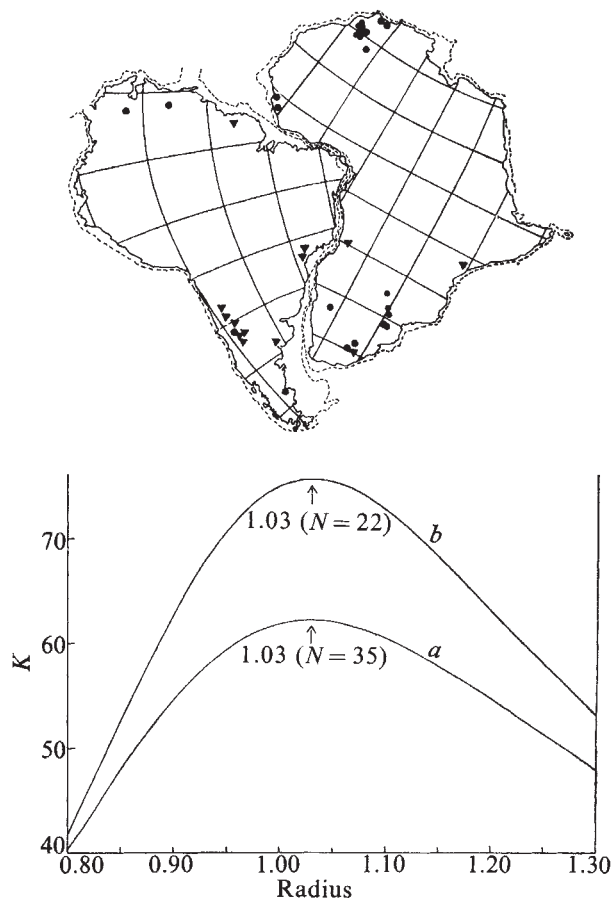
A significant achievement of the Apollo Programme was the establishment of the absolute age of these events. This knowledge, obtainable only from the returned rock samples collected by the

astronauts, constitutes a major scientific justification for the manned lunar landings. The highland crust formed before 4,400 Myr and was subjected to a massive bombardment which terminated at 3,900 Myr (refs 33–35). The basalt lavas, generated by partial melting deep (100–400 km) in the lunar interior, poured across the highland surface for at least 700 Myr, beginning at or just before the termination of the bombardment of the highland crust (3,900 Myr) and continuing to at least 3,200 Myr (ref. 29). Possibly some late flows continued to about 2,600 Myr (ref. 32). The surface has remained essentially static since 3,200 Myr. Most of the geological record preserved on the Earth postdates these events. There is no evidence of large-scale expansion or contraction³⁶ or any displacements resembling those produced by plate motions on the Earth.

Some minor features of lunar topography call for comment. In the highlands, some alteration of crater shapes has occurred. Such occasional distortions are attributable to the degradation of craters during subsequent cratering or basin-forming events³⁷. These massive events transport debris widely across the lunar surface, with extensive formation of secondary craters³⁸, and crater chains. Many effects previously attributed to volcanic or tectonic activity on the lunar surface are now explained by this mechanism^{37,38}. These basin-forming events terminated at 3,900 Myr (ref. 33). Faint lineations, referred to as the 'lunar grid'³⁹ have been detected by some workers. These may represent a faint residual trace of an early crustal cooling history. The 'grid', however, may result principally from the overlapping of outthrown debris produced by the giant basin forming collisions. In any event, it was produced before 3,900 Myr.

On the mare surfaces, wrinkle ridges³⁶, concentrated around

Fig. 2 Calculation of the palaeoradius for late Permian to early Jurassic data from Africa and South America after reconstruction of the South Atlantic according to the fit of Bullard *et al.*²² *a*, Permo-Jurassic; *b*, Triassic-Jurassic. The variation of the Fisher¹⁹ precision parameter for the palaeomagnetic poles as a function of radius (past/present) is shown below the locality map (●, Triassic-Jurassic localities; ▲, Permo-Triassic localities).



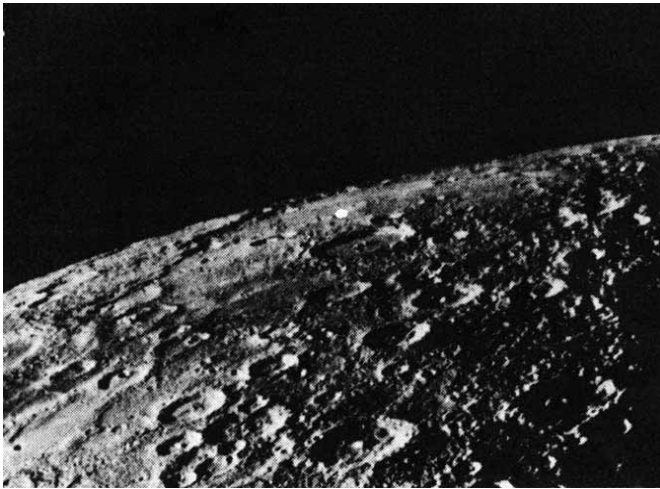


Fig. 3 Mariner 10 photograph of the northern limb of Mercury, showing a prominent east facing scarp, extending from the limb near the centre of the photograph, southwards for hundreds of kilometers. The linear dimension along the bottom of the photograph is 580 km. The picture was taken at a distance of 77,800 km from Mercury. (NASA Photograph P. 75-61-JPL-654-5-75).

the rims of the circular maria, seem to be due to localised cooling and draping of the lava sheets over pre-existing ridges. Occasional small grabens (straight rilles) in the maria are attributable to local cooling and shrinkage. Radial and concentric fractures in some basalt-filled craters²⁸ (for example, Gassendi, Humboldt) are apparently due to the same cause. The lack of present moonquake activity associated with these features is additional evidence for their antiquity. The most impressive features of the mare surface on a global scale are their smoothness (attributable to the low viscosity of the iron-rich lunar lavas) and the absence of evidence of tectonic movement since solidification.

The surface of the moon thus seems to have been frozen in its present form since about 3,200 Myr. It is possible to extrapolate this to 3,900 Myr when the last major structural events (formation of the Imbrium and Orientale ringed basins) occurred³³.

In contrast to the Moon, Mars shows evidence of slight expansion^{31,36,41}. Estimates of the chronology of the complex geological history discernible on Mars are not well established, being dependent on crater flux estimates and analogies with events on the moon^{42,43}. An early heavy bombardment produced the cratered terrain, still preserved in the southern hemisphere³¹. The next significant event seems to be the commencement of the uplift of the Tharsis region. This continent-size plateau is now about 6,000 km across and up to 10 km high. Radial fracturing of the martian crust accompanied this uplift. The next stage was the widespread production of volcanic plains, covering most of the northern hemisphere. A second stage of the Tharsis uplift now occurred, with further radial fracturing and the formation of large troughs. Finally the build-up of the large volcanic edifices (such as Olympus Mons) took place.

The dating of these events is difficult due to uncertainties in the crater flux rates. Present views^{42,43} date the old cratered terrain at about 4,000 Myr or older, from analogy with the lunar highlands.

The final uplift of the Tharsis plateau occurred about 1 Myr ago, and the large volcanoes are perhaps 200 Myr old^{42,43}. The Tharsis uplift thus seems to have occurred between about 4,000 and 1,000 Myr ago.

No major tensional features are observable on Mercury⁴³⁻⁴⁵. Extensive lobate scarps (Fig. 3) (20–500 km long and from 0.2–3 km high) provide evidence of contraction, with an upper limit of about 2 km since the formation of the crust. This effect is attributed to the contraction of a 600 km thick silicate shell around a cooling iron core⁴⁵. These scarps clearly formed later than the crust itself. From the observable crosscutting relationships, the formation of the scarps occurred close to the end of the period of major bombardment of the Mercurian surface, extending possibly a little beyond the stage^{44,45}. By analogy with the lunar highland surface, which the Mercurian landscape greatly resembles⁴⁴, the age of the heavily cratered surface is at least 4,000 Myr.

Table 3 provides estimates of the maximum change in planetary radii. For the moon, the expansion or contraction since 3,900 Myr is $< \pm 1$ km. For Mercury, a contraction of about 2 km has occurred probably before about 4,000 Myr (ref. 46). Since then, there is no evidence to support any further contraction or expansion. We therefore place limits of ± 1 km for the change in radius of Mercury over the past 3,000 to 4,000 Myr.

For Mars, the problem is more complex. The Tharsis Plateau is here assumed to be an uplifted feature due ultimately to thermal expansion^{31,36}, the case most favourable for variable G hypotheses. An alternative that it is a residual bulge following planetary contraction is less obviously reconcilable with the geomorphic evidence³¹. The maximum expansion has been estimated from thermal model calculations³⁶ to be about 19 km between 4,600 and 1,000 Myr. There seems to be no evidence for any expansion over the past 1,000 Myr. In Table 3, we give two estimates of the possible change in martian radius. Mars Model A assumes a 19 km expansion over 3,600 Myr, corresponding to the earlier martian record. Mars B assumes a possible ± 1 km change in radius during the past 1,000 Myr.

Variation of gravitational constant

Cosmologists have generally accepted the view that distant matter in the universe spherically distributed about the Earth has no noticeable effect on the Solar System. As the Universe expands and distant matter moves away from us it is supposed that there are no locally induced consequences. Exceptions to this view have been propounded by Dirac⁵, Jordan⁶, Brans and Dicke⁷, and Hoyle and Narlikar⁸. Jordan⁴⁷ later adopted a new cosmology closely paralleling that of Brans and Dicke. If the expanding Universe results in changing effects on distant matter, the consequences are a steadily decreasing gravitational constant. The Brans–Dicke theory of gravitation proposes a decrease in G of about $1 \times 10^{-11} \text{ yr}^{-1}$, for $\omega = 6$, where ω is the scalar coupling constant⁴⁸. But, recent lunar laser ranging data⁴⁹ require $\omega > 29$, and this limits the decrease in G to about $2 \times 10^{-12} \text{ yr}^{-1}$ for this theory.

In contrast, the theories of Dirac⁵² and Hoyle and Narlikar⁸ predict decreases of about $5 \times 10^{-11} \text{ yr}^{-1}$, when G is measured in atomic units. In Dirac's theory, there is also continuous creation of matter, and Dirac proposed two possibilities: additive creation and multiplicative creation⁵². In additive creation, matter is

Table 3 Palaeoradii (R_a) given as a function of present radius for the Earth, Moon, Mars and Mercury

Planet	R_a	α	Time (10^9 yr)	$-\delta G/G \text{ yr}^{-1}$	
				Constant mass	Multiplicative creation
Earth	1.020 ± 0.028	0.085 ± 0.02	0.4	$\leq 3 \times 10^{-10}$	$\leq 5 \times 10^{-9}$
Moon	1.0000 ± 0.0006	0.0004 ± 0.001	3.9	$\leq 5 \times 10^{-11}$	$\leq 1.5 \times 10^{-10}$
Mars A	0.9944	0.03 ± 0.01	3.6	$\leq 8 \times 10^{-11}$	—
Mars B	1.0000 ± 0.0003	0.03 ± 0.01	1.0	$\leq 1.5 \times 10^{-11}$	$\leq 4.5 \times 10^{-11}$
Mercury	1.0000 ± 0.0004	0.02 ± 0.005	3.5	$\leq 8 \times 10^{-12}$	$\leq 2.5 \times 10^{-11}$

The theoretical parameter α (see equation (3)) is listed and the corresponding maximum limits to the rate of decrease of the gravitational constant G are indicated. The models corresponding to Mars A and B are explained in the text.

created uniformly throughout the Universe and only the variation of G affects the planetary structure. In multiplicative creation, matter is created locally in proportion to the matter already present. In this case, the mass of a planet M increases at twice the rate as G decreases.

If the planet is in hydrostatic equilibrium then the pressure P at radius r within the planet satisfies

$$\frac{dP}{dr} = -\frac{G\rho(r)M(r)}{r^2} \quad (1)$$

where $M(r)$ is the mass within radius r . This can be rewritten as

$$\frac{dP}{dr^*} = -GM^{2/3}\frac{\rho(r^*)M^*(r^*)}{r^{*2}} \quad (2)$$

where $r^* \equiv r/M^{1/3}$, $M^*(r^*) = M(r)/M$. It follows that $R^* \equiv R/M^{1/3}$, where R is the radius of the planet, is a function only of $GM^{2/3}$ and the equation of state $P(\rho)$. A homologous change ($R \propto M^{1/3}$) would not be detectable in the observations we consider, since it scales all linear dimensions equally, so only variations in R^* need to be considered. In Dirac's Multiplicative Creation (DMC), $G \propto t^{-1}$ and $M \propto t^2$, where t is the elapsed atomic time, so $GM^{2/3} \propto t^{1/3} \propto G^{-1/3}$. It follows that for our considerations, DMC is exactly equivalent to a constant-mass theory in which G increases at one third of its actual rate of decrease. We can, therefore, restrict the following analysis to constant M .

The variation of R as G changes can be parameterised as

$$\frac{1}{R} \frac{dR}{dt} = -\frac{\alpha}{G} \frac{dG}{dt} \quad (3)$$

where α is to be calculated as a function of G and M . For a given $P(\rho)$ and the boundary condition $P(r=R)=0$, the solution of equation (1) for different values of G enables one to calculate α . For the particularly simple case of a polytropic gas, $P = C\rho^n$, dimensional considerations alone demonstrate that $\alpha = 1/(3n-4)$, independent of G and M . This result was applied by Hoyle and Narlikar⁵³ to the Earth, and by Crossley and Stevens⁵⁴ to Mercury, but neither application is valid as $\rho = 0$ at $P = 0$ in a polytropic gas, whereas the surface density of the terrestrial planets is finite. Because the value of α for small planets is required, it is more appropriate to derive a theory for the limit of small gravitational self-compression. This is certainly valid for the Moon, Mercury and Mars; and can be extended to include the Earth. Neglecting phase transitions and chemical layering for the moment, it is sufficient in this limit to set $P = K_0(\rho - \rho_0)/\rho_0$ in equation (1), where ρ_0 and K_0 are the zero-pressure density and incompressibility. Solution of equation (1) to lowest nonvanishing order in G is then:

$$\rho = \rho_0[1 + (\Delta\rho/\rho_0)(1 - r^2/R^2)]$$

$$\Delta\rho/\rho_0 \equiv 2\pi G\rho_0^2 R^2/3K_0 \quad (4)$$

The total mass M can be evaluated. The requirement that M be constant, as G and R change, gives

$$\alpha = \frac{2}{15} \left(\frac{\Delta\rho}{\rho_0} \right) \quad (5)$$

where $\Delta\rho$ is the density difference between planetary centre and surface caused by gravitational self-compression. A better approximation for α can be found by replacing ρ_0 by the average density $\bar{\rho}$ and K_0 by the average incompressibility $\bar{K}$. If the planet consists of chemical layers (each of which has constant mass as G changes) then $\Delta\rho/\bar{\rho}$ can be replaced by $\Sigma\Delta\rho_i/\bar{\rho}_i$, where $\Delta\rho_i$ is the density contrast across layer i of average density $\bar{\rho}_i$.

Univariant phase transitions (such as the olivine-spinel transition in the Earth and in Mars) cannot be treated this way, as the mass in each layer changes as G changes. For a homogeneous planet, the contribution to α because of a phase transition is

$$\alpha_{ph} = \frac{(\rho_1 - \rho_2)R_1(R^2 - R_1^2)}{2\rho_2 R^3} \quad (6)$$

in the limit of small self-gravitation, where $\rho = \rho_1$ for $0 \leq r \leq R_1$ and $\rho = \rho_2$ for $R_1 \leq r \leq R$.

For the Earth, equation (5) is a rather poor approximation and an extension of Birch analysis⁵⁵, including the olivine-spinel and spinel-postspinel transitions⁵⁶, gives $\alpha = 0.085 \pm 0.02$.

For Mars, the models of Okal and Anderson⁵⁷ (including the olivine-spinel transition) give $\alpha = 0.032$, whereas models incorporating a high pressure phase of magnetite⁵⁸ predict $\alpha \approx 0.028$. The martian moment of inertia is still uncertain⁵⁹, but $\alpha = 0.03 \pm 0.01$ encompasses all likely models.

Since Mercury provides the best constraint on the changes in R and G , the calculation for α has been done by both the simple method (equation (5)) and by the solution of equation (1) for several closely spaced values of G . For the differentiated models of Siegfried and Solomon⁶⁰ and their equations of state, the methods agree to within 20% and imply $\alpha = 0.02 \pm 0.005$. (The error is intended to indicate the likely uncertainty in the interior models of Mercury, given that the average density implies a predominantly iron composition).

For the Moon, any acceptable iron core has a negligible effect, and equation (5) applied to Ringwoods' models⁶¹ gives $\alpha = 0.004 \pm 0.001$.

The palaeoradii and these estimates for α give the upper limits to the rate of decrease of G listed in Table 3. Note that the limits for DMC are based on the limits for contraction of the planets, in accordance with our earlier discussion. To assess the validity of these limits, it is necessary to examine critically the assumption of hydrostatic equilibrium. Suppose that only an outer rigid lithosphere of thickness d violates hydrostatic equilibrium. As G decreases at constant R , the tensile stress in this lithosphere is about $3(\Delta G/G)\alpha K_0 R/d$. For $\Delta G = 0.03G$ (the upper limit implied by constant-mass Mercury), this stress exceeds 10 kbar, provided d is less than about 1,000 km (Earth, Mars), 500 km (Mercury), and 100 km (Moon). Since the lunar lithosphere is almost certainly much thicker than 100 km⁶², the allowable change in g might not affect the lunar surface. In contrast, the Mercurian mantle is only about 650 km thick⁶⁰, and the lithosphere is much thinner than this, as the presence of a large magnetic field suggests a thermally active planet with a large iron core^{63,64}.

If we exclude a fortuitous cancellation of thermal contraction and the expansion resulting from decreasing G in Mercury, then the upper limit of $8 \times 10^{-12} \text{ yr}^{-1}$ for the rate at which G decreases in a constant mass theory is realistic. This would seem to rule out the Dirac additive creation and Hoyle-Narlikar theories. The limit indicated by palaeomagnetic measurements on Earth is similar to that found from an analysis of radar observations in the inner Solar System^{65,66} and approaches the values predicted by the Dirac additive creation and Hoyle-Narlikar theories.

The upper limit for Dirac multiplicative creation of $2.5 \times 10^{-11} \text{ yr}^{-1}$ is a factor of 2 smaller than the value of $5 \times 10^{-11} \text{ yr}^{-1}$ indicated by the presently favoured value of Hubble's constant. This is a stronger constraint on the theory than other entirely different considerations⁶⁷⁻⁶⁹ have imposed.

Received 13 July; accepted 3 November 1977.

- Carey, S. W. *Earth Sci. Rev.* **11**, 105 (1975).
- Carey, S. W. *The Expanding Earth* (Elsevier, Amsterdam, 1977).
- Owen, H. G. *Phil. Trans. R. Soc. A* **281**, 223 (1976).
- Steiner, J. *Geology* **5**, 313 (1977).
- Dirac, P. A. M. *Proc. R. Soc. A* **165**, 199 (1938).
- Jordan, P. *Nature* **164**, 637 (1949).
- Brans, C. & Dicke, R. H. *Phys. Rev.* **124**, 925 (1961).
- Hoyle, F. & Narlikar, J. V. *Nature* **233**, 41 (1971).
- Van Andel, S. I. & Hospers, J. *Tectonophysics* **6**, 491 (1968).
- Egyed, L. *Geofis. Para Appl.* **45**, 115 (1960).
- Egyed, L. *Nature* **190**, 1097 (1961).
- Van Hilten, D. *Nature* **200**, 1277 (1963).
- Hospers, J. & Van Andel, S. I. *Tectonophysics* **5**, 5 (1967).

14. Van Andel, S. I. & Hospers, J. *Tectonophysics* **5**, 273 (1968).
15. Carey, S. W. in *Proc. Symp. Continental Drift* 177 (1958).
16. Hilgenberg, O. *Neu. Jahrb. Geol. Palaeontol.* **116**, 1 (1962).
17. Ward, M. A. *Geophys. J. R. astr. Soc.* **8**, 217 (1963).
18. Ward, M. A. *Geophys. J. R. astr. Soc.* **10**, 445 (1966).
19. Fisher, R. A. *Proc. R. Soc. A* **217**, 295 (1953).
20. McElhinny, M. W. *Palaeomagnetism and Plate Tectonics* (Cambridge University Press, Cambridge, 1973).
21. McElhinny, M. W. & Brock, A. *Earth planet. Sci. Lett.* **27**, 321 (1975).
22. Bullard, E. C., Everett, J. E. & Smith, A. G. *Phil. Trans. R. Soc.* **258**, 41 (1965).
23. Smith, A. G. & Hallam, A. *Nature* **225**, 139 (1970).
24. Eged, L. *Nature* **197**, 1059 (1963).
25. Wesson, P. S. *Q. J. R. astr. Soc.* **14**, 9 (1973).
26. Lyell, C. *Principles of Geology* 1 (John Murray, London, 1830).
27. Tennyson, A. *Lord In Memoriam* Section 123 (1850).
28. Mutch, T. A., *Geology of the Moon: A Stratigraphic View* (Princeton University Press, Princeton, 1972).
29. Taylor, S. R. *Lunar Science: A Post-Apollo View* (Pergamon, New York, 1975).
30. Gault, D. E., Guest, J. E., Murray, J. B., Dzuris, D. & Malin, M. C. *J. geophys. Res.* **80**, 2444 (1975).
31. Mutch, T. A., Arvidson, R. E., Head, J. W., Jones, K. L. & Saunders, R. S. *The Geology of Mars* (Princeton University Press, Princeton, 1976).
32. Head, J. W. *Rev. geophys. Space Phys.* **14**, 265 (1976).
33. Tera, F., Papanastassiou, D. A. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **22**, 1 (1974).
34. Schaeffer, O. A., Husain, L. & Schaeffer, G. A. *Proc. 7th Lunar Sci. Conf.* 2067 (1976).
35. Turner, G. & Cadogan, P. H. *Proc. 6th Lunar Sci. Conf.* 1509 (1975).
36. Solomon, S. C. & Chaiken, J. *Proc. 7th Lunar Sci. Conf.* 3229 (1976).
37. Wilhems, D. E., *Proc. 7th Lunar Sci. Conf.* 2883 (1976).
38. Oberbeck, V. R. *Rev. geophys. Space Phys.* **13**, 337 (1975).
39. Fielder, G. *Structure of the Moon's Surface* (Pergamon, London, 1961).
40. Bryan, W. B. *Proc. 4th Lunar Sci. Conf.* 93 (1973).
41. Blasius, K. R. & Cutts, J. A. *Proc. 7th Lunar Sci. Conf.* 3561 (1976).
42. Soderblom, L. A., West, R. A., Herman, B. M., Kriedler, T. J. & Condit, C. D. *Icarus* **22**, 239 (1974).
43. Carr, M. J. *geophys. Res.* **79**, 3943 (1974).
44. Stromm, R. G., Trask, N. J. & Guest, J. E. *J. geophys. Res.* **80**, 2478 (1975).
45. Murray, B. C., Stromm, R. G., Trask, N. J. & Gault, D. E. *J. geophys. Res.* **80**, 2508 (1975).
46. Solomon, S. C. *Icarus* **28**, 509 (1976).
47. Jordan, P. Z. *Astrophys. J.* **68**, 201 (1968).
48. Dicke, R. H. *Science* **138**, 653 (1962).
49. Williams, J. G. *et al. Phys. Rev. Lett.* **36**, 551 (1976).
50. Lytle, R. A. *The Moon* **16**, 41 (1976).
51. Jacobs, J. A. *The Earth's Core* 187 (Academic, New York, 1975).
52. Dirac, P. A. M. *Proc. R. Soc. A* **338**, 439 (1974).
53. Hoyle, F. & Narlikar, J. V. *Mon. not. R. astr. Soc.* **155**, 323 (1972).
54. Crossley, D. J. & Stevens, R. K. *Can. J. Earth Sci.* **13**, 1723 (1976).
55. Birch, F. *Phys. Earth planet. Int.* **1**, 141 (1968).
56. Ringwood, A. E. *Composition and Petrology of the Earth's Mantle* (McGraw-Hill, New York, 1975).
57. Okal, E. A. & Anderson, D. L. *Icarus* (submitted).
58. Ringwood, A. E. & Clark, S. P. *Nature* **234**, 89 (1971).
59. Reasenberg, R. D. *J. geophys. Res.* **82**, 369 (1977).
60. Siegfried, R. W., II & Solomon, S. C. *Icarus* **23**, 192 (1974).
61. Ringwood, A. E. *Phil. Trans. R. Soc.* **285**, 577 (1977).
62. Kuckes, A. F. *Phys. Earth planet. Int.* **14**, 1 (1977).
63. Stevenson, D. J. *Nature* **256**, 634 (1975).
64. Cassen, P., Young, R. E., Schubert, G. & Reynolds, R. T. *Icarus* **28**, 501 (1976).
65. Shapiro, I. I., Smith, W. B., Ash, M. B., Ingalls, R. P. & Pettergill, G. H. *Phys. Rev. Lett.* **26**, 27 (1971).
66. Reasenberg, R. D., Shapiro, I. I., Pettengill, G. H. & Campbell, D. B. *Bull. Am. astr. Soc.* **8**, 308 (1976).
67. Canuto, V., Adams, P. J. & Tsiang, E. *Nature* **261**, 438 (1976).
68. Macder, A. *Astr. Astrophys.* **57**, 125 (1977).
69. Roxburgh, I. W. *Nature* **268**, 504 (1977).

West Antarctic ice sheet and CO₂ greenhouse effect: a threat of disaster

J. H. Mercer

Institute of Polar Studies, The Ohio State University, Columbus, Ohio 43210

If the global consumption of fossil fuels continues to grow at its present rate, atmospheric CO₂ content will double in about 50 years. Climatic models suggest that the resultant greenhouse-warming effect will be greatly magnified in high latitudes. The computed temperature rise at lat 80° S could start rapid deglaciation of West Antarctica, leading to a 5 m rise in sea level.

ATMOSPHERIC carbon dioxide traps some of the long-wave radiation emitted by the Earth's surface (principally near 15 μ m wavelength), thereby tending to warm the troposphere. This so-called greenhouse effect has long been suspected^{1,2} but only recently, as the implications of a continuation of the current near-exponential growth of industrial CO₂ production have been realised, have many come to fear a disastrous climatic warming in the rather near future. In a recent report on the climatic effects of energy production, Revelle *et al.*³ conclude that industrial civilisation may soon have to decide whether or not to make the tremendous investment of capital and effort needed to change over from fossil fuels to other sources of energy. Bolin⁴, in hearings before the U.S. House of Representatives, points out that if all reserves of fossil fuel that are accessible by present techniques were burnt, thereby increasing atmospheric CO₂ content between five- and eightfold above its preindustrial level, the result would almost certainly be climatic disaster.

I contend that a major disaster—a rapid 5 m rise in sea level, caused by deglaciation of West Antarctica—may be imminent or in progress after atmospheric CO₂ content has only doubled. This concentration of CO₂ will be reached within about 50 years if fossil fuel continues to be consumed at its recent accelerating rate, or within about 200 years if consumption is held constant at today's level^{4,5}. Keeling and Bacastow⁶ believe that even if a policy of conversion to other sources of energy was started today and vigorously pursued, the global rate of consumption of fossil fuels would still double by the end of this century.

If so, the actual doubling time for atmospheric CO₂ content is likely to be nearer 50 than 200 years.

Many attempts have been made to estimate by climatic modelling the average global rise in temperature that would result from a doubling of atmosphere CO₂ content. The figures obtained have ranged from 0.7 K to 9.6 K, and Schneider⁷ has critically examined the models in an attempt to clear up the confusion created by these widely different estimates. He points out that some of the models give unrealistic results because they compute an equilibrium condition for the Earth's surface rather than for the Earth-atmosphere system as a whole. He stresses the advantages of radiative-convective models, which take into account vertical motions of the atmosphere and latent heat transport, and he compares the radiative-convective models of Rasool and Schneider⁸, who had computed an average global temperature rise of 0.8 K, with that of Manabe and Wetherald⁹ who had computed a rise of 2.3 K, later revising this to 2.9 K. He estimates that, using the most refined input of feedback mechanisms that is possible with present knowledge, globally-averaged temperatures would rise about 1.9 K. But, because some feedback mechanisms may have been improperly modelled, some (especially those involving changes in cloud cover and cloud top elevation) have been largely ignored because of our inability to do otherwise, and others perhaps remain unknown, he concludes that the best state of the art, order of magnitude estimate is that a doubling of atmospheric CO₂ content would raise average global temperature by 1.5–3 K. He believes that this estimate is as likely to be too low as too high.

More recently Augustsson and Ramanathan¹⁰, using a different modelling technique, and taking into account the previously ignored effect of bands of weak absorption by CO₂, have calculated that global temperatures would rise about 2 K. This, they note, is similar to Schneider's figure; they find this encouraging, but surprising and perhaps fortuitous because substantially different modelling techniques were used. Their one-dimensional model gives no information about differences between latitudinal zones, but the three-dimensional model of

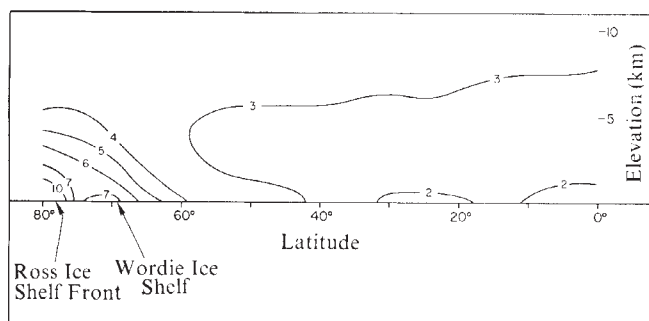


Fig. 1 Computed rise in mean atmospheric temperature, as a function of latitude and elevation, that would result from a doubling of atmospheric CO_2 content (after Manabe and Wetherald⁹ who emphasise that, because of the various simplifications of the model, the quantitative aspects of the results should not be taken too seriously).

Manabe and Wetherald⁹ shows greatly magnified warming in high latitudes caused by the lack of vertical mixing, and by the feedback effect of decreased albedo as snow and sea ice cover recede: temperatures would rise by ~ 2 K between the equator and lat 30° north or south, by 4 K at lat 60° , by 7 K at lat 70° , and by > 10 K at lat 80° (Fig. 1). Schneider⁷ agrees that this effect is likely and points out that warming in the polar regions could be crucial because of their sensitivity to changes in the energy balance.

Manabe and Wetherald⁹ emphasise that because their model is highly simplified—it has idealised global topography, fixed cloudiness, no heat transport by ocean currents and no seasonal variability—these figures for the rise in temperature should not be taken at their face value. Smagorinsky¹¹ discusses the model and points out that seasonal variability is of fundamental importance in high latitudes because it determines the extent of snow and ice cover; if this factor, and the reactions of the oceans and clouds were taken into account, a very different result might be obtained; it is even possible that global cooling would be indicated. Keeling and Bacastow⁶ conclude that until the feedback effect of the slow response of subsurface ocean waters can be correctly modelled, the regional climatic changes that are of greatest interest to mankind will be hard to predict. Nevertheless, as Bolin⁴ points out, although the model of Manabe and Wetherald has serious shortcomings it is the most advanced that has yet been developed, and to gamble with the Earth's climate by ignoring it would be highly irresponsible. In the same vein Schneider¹² sums up the dilemma facing mankind: despite the crudities and inadequacies of present techniques for modelling the climatic effects of increasing atmospheric CO_2 content and the resultant doubts about the magnitude of the warming that would actually occur, we cannot afford to let the atmosphere carry out the experiment before taking action because if the results confirm the prognosis, and we should know one way or the other by the end of the century, it will be too late to remedy the situation on account of the long residence time of CO_2 in the atmosphere (Keeling and Bacastow⁶ estimate that, if all accessible fossil fuels were burnt, restoration of pre-industrial levels of CO_2 would take at least 10,000 yr).

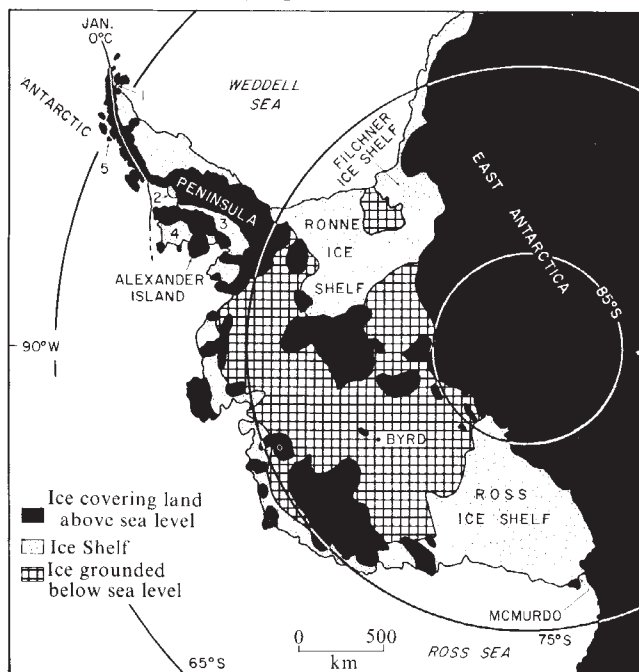
Is the CO_2 greenhouse effect detectable in recent climatic trends?

Since about 1940, temperatures over much of the Northern Hemisphere have dropped despite rising atmospheric CO_2 content. Broecker¹³ suggests that this may have lulled us into a sense of false security about the danger of increasing CO_2 levels; the cooling does not disprove or cast doubt on the CO_2 greenhouse-warming theory but is, he suspects, merely the effect of a natural cooling cycle that has overwhelmed the warming due to CO_2 , which he estimates has amounted to no more than 0.2°C for the past 30 years. He

believes that the quasi-cyclical pattern of climatic fluctuations in the recent past, which is shown by the oxygen isotope record from the Greenland Ice Sheet¹⁴, implies that this cooling will soon level out, to be followed by a period of rapid warming as the natural climatic trend is reinforced by the effects of increasing atmospheric CO_2 . In fact, the cooling since 1940 seems to have been mainly confined to middle and high latitudes in the Northern Hemisphere, and some investigators believe that the southern part of the Southern Hemisphere has warmed during the same interval. Damon and Kunen¹⁵ have studied climatic records from 67 Southern Hemisphere stations that meet certain specifications and that have records that go back to 1954 or earlier. They find that since 1943 temperatures have changed little between the equator and lat 45°S , except in Australia and New Zealand which, as other workers also point out^{16,17}, have warmed by about 1°C since the 1940s. South of lat 45°S , however, they conclude that average annual temperatures increased between the 1960–64 and 1970–74 pentads, particularly in West Antarctica where they rose about 2°C at Argentine Island (lat 65°S), McMurdo (lat 78°S), and Byrd (lat 80°S) (Fig 2). Thomas^{18,19} observation that temperatures at 10 m depth in the eastern part of the Ross Ice Shelf rose about 1°C between 1958 and 1972 confirms the warming trend in West Antarctica.

Damon and Kunen¹⁵ suggest that the climate of the Southern Hemisphere is now responding to the CO_2 greenhouse effect, whereas in the Northern Hemisphere this has recently been overwhelmed by cooling caused by man-made particulate pollution and also, possibly, from greater volcanic activity. Budd¹⁹, however, while noting the recent warming in West Antarctica, finds no clear evidence that Antarctica as a whole either warmed or cooled between 1958 and 1972; furthermore, at the South Pole 1976 was the coldest year since records began there in 1957²⁰. So far, apparently, despite recent regional warming over Australasia and West Antarctica, there is no unequivocal evidence for a global rise in temperature caused by increasing atmospheric CO_2 content. If Broecker's estimate is correct that average global CO_2 -caused warming would have amounted to no more than a fifth of a degree in the past 30 years, this is perhaps not surprising.

Fig. 2 West Antarctica, showing ice shelves, ice grounded below sea level, ice covering land above sea level, and position of the 0°C January isotherm in the Antarctic Peninsula (based on information up to the year 1962)²⁰. 1, Prince Gustav Channel; 2, Wordie Ice Shelf; 3, George VI Sound; 4, Wilkins Sound; 5, Argentine Island.



First disastrous consequence of rising CO₂ levels

Those who are most aware of the climatic dangers of increasing atmospheric CO₂ do not seem to view deglaciation in Antarctica as an immediate threat. For example, Schneider and Dickinson²¹, and Bolin⁴, believe that the Antarctic Ice Sheet will respond to climatic warming very slowly, during millennia, while Rotty and Budyko^{22,23} are more concerned with the shrinkage and eventual disappearance of the Arctic sea ice. Revelle *et al.*³ conclude that our present understanding of the Antarctic Ice Sheet is insufficient for us to forecast how it would be affected by global warming of several degrees. They believe that the direct effect of the warming would be minimal, but suggest that increased snowfall might eventually thicken the ice sheet enough to cause surging, perhaps resulting in deglaciation of West Antarctica. But, not only would effective thickening of the ice sheet require a long interval of increased snowfall, but also it is doubtful that a thickening ice sheet would surge. I believe that Revelle *et al.*³ underestimate the direct effect of moderate Antarctic warming, which would be likely to cause deglaciation of West Antarctica long before any significant thickening of the ice sheet could occur.

If present trends in fossil fuel consumption continue, and if the greenhouse warming effect of the resultant increasing atmospheric CO₂ is as great as the most advanced current models suggest, a critical level of warmth will have been passed in high southern latitudes 50 years from now, and deglaciation of West Antarctica will be imminent or in progress. Deglaciation would probably be rapid once it had started, and when complete would have led to a rise in sea level of about 5 m along most coasts. The reasons for this ominous situation lie in the unique characteristics of the West Antarctic ice sheet.

West Antarctic ice sheet's vulnerability to climatic warming

The Antarctic Ice Sheet consists of two unequal parts, with different histories and characteristics: the vast, long-established, mainly land-based ice sheet in East Antarctica, and the younger, much smaller marine ice sheet²⁴, grounded as much as 2,500 m below sea level in West Antarctica (Figs 2 and 3). Melting of the East Antarctic ice sheet would raise sea level by about 50 m, whereas melting of the West Antarctic ice sheet, much of which is only displacing ocean water, would raise sea level by no more than about 5 m on average²⁵ (Clark & Lingle²⁶ show that the sea level change would not be globally uniform).

When the Antarctic Ice Sheet formed while temperatures dropped during the late Cainozoic, there was an interval, during which temperatures dropped further, between the emplacement of the East Antarctic ice sheet in the late Miocene²⁷ and of the West Antarctic ice sheet, probably during the latest Miocene or earliest Pliocene²⁸. This was because the land-based East Antarctic ice sheet could form as a temperate glacier, whereas the marine West Antarctic ice sheet had to consist of cold ice from the start²⁵. (The thermal requirements of the West Antarctic ice sheet are discussed in more detail later.) Thus, in order to survive, the West Antarctic ice sheet needs colder summers—perhaps as much as 10 °C colder—than does the East Antarctic ice sheet, and is, therefore, more vulnerable to a rise in temperature. During sustained climatic warming, the Miocene–Pliocene sequence of glacial buildup would be reversed, and the West Antarctic ice sheet would be eliminated before the East Antarctic ice sheet was greatly affected. Fortunately, serious depletion of the East Antarctic ice sheet is a distant threat because, being land-based, it could retain a positive mass balance after climatic warming had converted it from a cold to a temperate glacier at low elevations. Even if further warming gave it a severely negative mass balance, it would waste away only slowly over millennia, by *in situ* melting, as the Laurentide Ice Sheet did 14–8,000 yr ago²⁹. The resultant rise

in sea level would cause considerable inconvenience, but would be gradual enough to allow coastal communities to adjust. This is apparently the type of deglaciation that Schneider and Dickinson²¹ and Bolin⁴ had in mind when they considered the effects of climatic warming on polar ice sheets; unfortunately, however, the West Antarctic ice sheet would be unlikely to shrink in this manner.

In contrast to the land-based East Antarctic ice sheet, the marine ice sheet in West Antarctica can exist only so long as its grounded portion is buttressed by fringing ice shelves; in particular, by the Ross and Filchner–Ronne ice shelves (Fig. 2). Thus any environmental change that diminished or destroyed these ice shelves would also diminish or destroy the ice grounded below sea level; as the ice shelf fronts receded southward, their grounding lines would also recede and the grounded ice sheet would shrink and thin. Eventually, after all ice shelves had disappeared, ice cover would be confined to areas above sea level, and glaciers would terminate in ice cliffs between high and low water levels^{25,30–32}.

Ice shelves are vulnerable to both oceanic and atmospheric warming. They will melt at the base if the water they are in contact with is above its freezing point; as Robin³⁴ notes, they are absent from the coasts of the Antarctic Peninsula only where sea temperatures rise above –1.5 °C during the warmest month. (For water with 35‰ salinity, freezing point is about –1.9 °C at the surface and about –2.2 °C at 500 m depth³³). Melting will accelerate if a positive temperature gradient develops between the base and the surface; this will happen only if the ice shelf becomes temperate as rising air temperatures produce enough meltwater to percolate downwards and destroy the previous winter's cold wave. In fact temperate ice shelves do not seem to exist in nature, except possibly as short-lived features during climatic warming; Robin and Adie³¹ observe that all known ice shelves are cold, that is, below the pressure melting point at depth, and that their northern limit on the west coast of the Antarctic Peninsula lies a short distance south of the northernmost land glaciers that are cold at sea level. They conclude that a climatic warming above a critical level would remove all ice shelves and, consequently, all ice grounded below sea level, resulting in the deglaciation of most of West Antarctica.

The observations of Robin and Adie imply that where summers are warm enough to destroy the previous winter's cold wave in a glacier at sea level, ice shelves will be absent. In the Antarctic Peninsula the 0 °C isotherm for the warmest month (January), which trends south-west from the tip of the peninsula towards Alexander Island³⁵ (Fig. 2), is almost parallel to, and a short distance outside, the northern limit of ice shelves. Other aspects of the summer climate besides air temperature, particularly the duration and intensity of solar radiation, must also be involved in determining the northern limit of cold ice at sea level, and thus of the ice shelves, but the 0 °C isotherm for midsummer air temperature seems at least as realistic a climatic boundary for ice shelves as does, for example, the 10 °C midsummer isotherm that approximately marks the northern limit of trees³⁶. Most ice shelves, however, terminate considerably south of the 0 °C midsummer isotherm. Around the coasts of East Antarctica they are limited by the positions of the outermost lateral anchor points³⁷, but in West Antarctica neither high temperatures nor lack of anchor points prevent the northward expansion of the Ross and Filchner–Ronne ice shelves. The frontal positions of these large ice shelves depend mainly on the position of the grounding line of the West Antarctic ice sheet, and this is determined by water depth and, to a smaller extent, by snow accumulation rates.

Average midsummer air temperatures at the fronts of both the Ross and Filchner–Ronne ice shelves are now about

4 °C to –5 °C³⁵. Thus the deglaciation of West Antarctica that Robin and Adie³¹ foresaw would result from sufficient climatic warming would be unlikely to start until air temperatures had risen about 5 °C, bringing summer temperatures

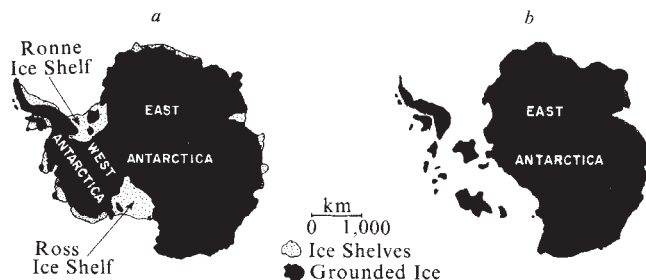


Fig. 3 a, Antarctic ice cover today, and b, after a 5–10 °C warming.

over the Ross and Filchner–Ronne ice shelves up to those now prevailing in the northwestern Antarctic Peninsula. Once this level of comparative warmth had been reached, deglaciation would probably be rapid, perhaps catastrophically so, because even a small additional rise in temperature would affect a large expanse of gently sloping ice shelf. As Hughes³² graphically describes it, “a relatively minor climatic fluctuation along the ice shelf calving barrier can unleash dynamic processes independent of climate that cause calving bays to remorselessly carve out the living heart of a marine ice sheet.” An example of deglaciation by the calving bay mechanism was the extremely rapid disintegration of the central portion of the Laurentide Ice Sheet, centred over the present Hudson Bay, sometime between 8,000 and 7,500 BP, after thousands of years of slow recession on land³⁸.

The Ross and Filchner–Ronne ice shelves would probably be on the verge of receding southward if summer temperatures rose 5 °C at lat 80° S. The climatic model of Manabe and Wetherald⁹, for which year-round climatic uniformity is assumed, suggests a warming of more than 10 °C at lat 80° S after atmospheric CO₂ content has doubled; this concentration will perhaps be reached within 50 years. If this model is even approximately correct, rather drastic deglaciation should then be in progress in West Antarctica.

One warning sign that a dangerous warming is beginning in Antarctica will be the breakup of ice shelves in the Antarctic Peninsula just south of the recent January 0 °C isotherm; the ice shelf in Prince Gustav Channel on the east side of the peninsula, and the Wordie Ice Shelf, the ice shelf in George VI Sound, and the ice shelf in Wilkins Sound on the west side (Fig. 2). There is some evidence, not yet conclusive, that such a southward recession of the ice shelf limit in the Antarctic Peninsula has already begun; a comparison of LANDSAT imagery with earlier surveys has shown that between 1966 and 1974 nearly 600 km² or about one quarter of the Wordie Ice Shelf broke away, and the ice shelf in George VI Sound receded³⁹.

Fall in sea level by glacial growth

High latitude warming might well bring increased snowfall to Antarctica and perhaps to parts of Greenland also, as Revelle *et al.*³ suggest. This is far from certain, however, because although the warmer air masses could hold more moisture, cyclonic disturbances might decrease in number and intensity if, as present models indicate, the latitudinal temperature gradient decreased. To what extent deglaciation in West Antarctica would be offset by glacial buildup elsewhere is hard to estimate. Robin and Adie³¹ calculate that doubling the net annual accumulation rate on the East Antarctic ice sheet, which already reaches the coast, would increase its volume by no more than 10%. This, they estimate, would eventually result in a 4 m drop in sea level; however, they also conclude that the East Antarctic ice sheet would take several thousand years to reach its new equilibrium volume. Thus, if temperatures rose no further, warmth-induced deglaciation of West Antarctica might eventually be partially or even wholly offset

by thickening of the surviving ice sheets. It would be unwise, however, to rely on this buildup being rapid enough to avert a disastrous rise in sea level, and in any case, if temperatures continued to rise because of increasing atmospheric CO₂ content, glacial growth in East Antarctica would eventually cease.

Previous destruction of the West Antarctic ice sheet

Considerable evidence from the Southern Hemisphere suggests that the warmest part of the last interglacial (Sangamon–Eem) was warmer than the present interglacial has been so far; for instance, subantarctic seas were then warmer than they have been since (J. D. Hays, personal communication), and in Southern Chile chemical weathering was unusually intense⁴⁰. This warm interval—substage 5e, according to the sequence established by Emiliani⁴¹—was centred 120–125,000 yr ago⁴². If the West Antarctic ice sheet was absent at that time, the hypothesis that a rather moderate rise in temperature would destroy it would be strengthened.

I earlier suggested that the high sea level of the last interglacial (probably about +6 m) resulted mainly from deglaciation of West Antarctica when temperatures there rose too high for the survival of ice shelves²⁸. Deglaciation of West Antarctica alone would add a layer of water about 5 m deep over the area of the present world ocean. Later, Emiliani suggested that the high sea level resulted from deglaciation of Greenland, with minimal contribution from Antarctica, but this is highly unlikely because the marine West Antarctic ice sheet is much more vulnerable to climatic warming than is the land-based Greenland ice sheet.^{43,44}

A rise in eustatic sea level is not necessarily a glacio–eustatic rise, so that the high sea level of the last interglacial by no means proves that less ice was then present. Supporting evidence for the hypothesis came later when oxygen isotopic analyses of Core V 28–238 from the equatorial Pacific showed that the mass of the oceans during substage 5e was greater than it is today⁴². The isotopic difference between the Holocene and substage 5e was about what might be expected from melting of the West Antarctic ice sheet during substage 5e. Shackleton (personal communication) now believes that because planktic and not benthic foraminifera were analysed for substage 5e in Core V28–238, the isotopic composition may include a temperature effect. Thus, although the oceans did contain more water during substage 5e than they do now, just how much more is uncertain. Measurements being made on other deep sea cores should give more accurate figures in the near future.

Possible disintegration of West Antarctic ice sheet from non-climatic causes

Some workers^{45–49} believe that the West Antarctic ice sheet shrank at the end of the last glaciation when its grounding line receded as the result of the ~120 m eustatic rise in sea level. Others^{50,51}, however, believe that changes in the extent and thickness of the West Antarctic ice sheet since the last glacial age have been minor. In any case, with the larger Northern Hemisphere ice sheets gone, no further major shrinkage of the West Antarctic ice sheet as the result of rising sea level will occur in the near future.

Thomas and Weertman^{52,53} believe that the present West Antarctic ice sheet is inherently unstable and that, if it disintegrates, it will do so through mechanisms unconnected with climate. Hughes⁵⁴ agrees that such internal instability mechanisms are possible and could result in surges, but he also stresses the vulnerability of the ice sheet to climatic change. Weertman⁵³ suggests that the presence of ice streams draining the ice sheet is a premonition of a large-scale surge, the start of a process that may cause the West Antarctic ice sheet to discharge one third to one half of its volume into the oceans over, say, 100 years. The only occasion, however, during the last half

million years or so that the oxygen isotopic composition of the oceans implies the presence of less land ice than today is during substage 5e (ref. 42); that is, during an interval that was warmer than the Holocene. This strongly suggests cause and effect: the West Antarctic ice sheet was absent because of the high temperature. There is no evidence that the ice sheet has ever been seriously depleted by surges in the past and therefore, I believe, there is no reason to suppose that it will be in the future. Weertman may be right for the wrong reasons; the ice sheet is likely to disintegrate in the rather near future, but because of man-made climatic warming, not through surges resulting from internal instability mechanisms.

Conclusions

The present West Antarctic ice sheet is reasonably secure against all instability mechanisms except for climatic warming above a critical and—for the Pleistocene—exceptional level. In the natural course of events, this level of warmth might be reached about once in half a million years, perhaps as a consequence of an unusual combination of the astronomic factors that seem to be responsible for the timing of the major glacial-interglacial climatic changes⁵⁵. Furthermore, because the present interglacial has apparently passed its natural peak of warmth—this was probably reached about 9,400 yr ago in the Southern Hemisphere⁵⁵—deglaciation of West Antarctica would not occur in the foreseeable future without Man's injection of massive amounts of industrial CO₂ into the atmosphere.

If the recent growth rate of fossil fuel consumption continues, atmospheric CO₂ content is expected to double in about 50 yr. Present models of the climatic effects of this doubling compute a rise in temperature that could cause rapid deglaciation of West Antarctica, leading to a 5 m rise in sea level. Although the models are known to be crude and over-simplified, so that the climatic changes that will actually occur will no doubt differ considerably from their estimates, there is, at present, no way of knowing whether the models err on the optimistic or pessimistic side.

If the CO₂ greenhouse effect is magnified in high latitudes, as now seems likely, deglaciation of West Antarctica would probably be the first disastrous result of continued fossil fuel consumption. A disquieting thought is that if the present highly simplified climatic models are even approximately correct, this deglaciation may be part of the price that must be paid in order to buy enough time for industrial civilisation to make the changeover from fossil fuels to other sources of energy. If so, major dislocations in coastal cities, and submergence of low-lying areas such as much of Florida and the Netherlands, lies ahead. More sophisticated climatic modelling may show that the outlook is less alarming than this, but on the other hand, it may show that the situation is even more threatening. The urgent need for this sophisticated modelling is evident.

One of the warning signs that a dangerous warming trend is under way in Antarctica will be the breakup of ice shelves

on both coasts of the Antarctic Peninsula, starting with the northernmost and extending gradually southward. These ice shelves should be regularly monitored by LANDSAT imagery.

I thank I. M. Whillans for helpful comments.

Received 1 August; accepted 3 November 1977.

1. Chamberlin, T. C. *J. Geol.* **7**, 547 (1899).
2. Plass, G. N. *Tellus* **8**, 140 (1956).
3. Revelle, R. R. and the Geophysics Study Committee in *Energy and Climate* (Studies in Geophysics) 1–40 (National Academy of Sciences, Washington, D.C., 1977).
4. Bolin, B. in *National Climate Program Act* 29–30 (U.S. Govt. Printing Office, Washington, D.C., 1976).
5. Hoffert, M. I. *Atmos. Environ.* **8**, 1225 (1974).
6. Keeling, C. D. & Bacastow, R. B. in *Energy and Climate* (Studies in Geophysics) 110–160 (National Academy of Sciences, Washington, D.C., 1977).
7. Schneider, S. H. *J. Atmos. Sci.* **32**, 2060 (1975).
8. Rasool, S. & Schneider, S. H. *Science* **173**, 135 (1971).
9. Manabe, S. & Wetherald, R. T. *J. Atmos. Sci.* **24**, 241 (1967); **32**, 3 (1975).
10. Augustsson, T. & Ramanathan, V. *J. Atmos. Sci.* **34**, 448 (1977).
11. Smagorinsky, J. in *Energy and Climate* (Studies in Geophysics) 229–242 (National Academy of Sciences, Washington, D.C., 1977).
12. Schneider, S. H. *The Genesis Strategy* 53 and 56 (Plenum, New York, 1976).
13. Broecker, W. S. *Science* **189**, 460 (1975).
14. Dansgaard, W., Johnsen, S. J., Clausen, H. B. & Langway, C. C. in *The Late Cenozoic Glacial Ages*, (ed. Turekian, K. K.) 37 (Yale University Press, New Haven, 1971).
15. Damon, P. E. & Kunen, S. M. *Science* **173**, 447 (1976).
16. Salinger, M. J. & Gunn, J. M. *Nature* **256**, 396 (1975).
17. Tucker, G. B. *Search* **6**, 323 (1975).
18. Thomas, R. H. *J. Glaciol.* **16**, 111 (1976).
19. Budd, W. F. *J. Glaciol.* **15**, 417 (1975).
20. *Antarctic J. U.S.* **12**, 51 (1977).
21. Schneider, S. H. & Dickinson, R. E. *Rev. Geophys. Space Phys.* **12**, 447 (1974).
22. Rotty, R. M. in *National Climate Program Act* 118 (U.S. Govt. Printing Office, Washington, D.C., 1976).
23. Budyko, M. I. *Trans. Am. Geophys. Union* **53**, 868 (1972).
24. Mercer, J. H. *Palaeogeogr., Palaeoclimat., Palaeoecol.* **8**, 21 (1970).
25. Mercer, J. H. in *Int. Ass. Sci. Hydrol. Commission of Snow and Ice, General Assembly of Bern, Publ. No. 79*, 217 (1968).
26. Clark, J. A. & Lingle, C. S. *Nature* **269**, 206 (1977).
27. Shackleton, N. J. & Kennett, J. P. in *Init. rep. Deep Sea Drilling Proj.* **29** (eds. Kennett, J. P., Houtz, R. E. et al.) 752 (U.S. Govt. Printing Office, Washington, D.C., 1975).
28. Mercer, J. H. in *Palaeo-ecology of Africa, and of the surrounding islands and Antarctica*, **8**, 85–114 (ed. van Zinderen Bakker, E. M.), Balkema, (Cape Town, 1975).
29. Bryson, R. A., Wendland, W. M., Ives, J. D. & Andrews, J. T. *Arctic and Alpine Res.* **1**, 1 (1969).
30. Hollin, J. T. *J. Glaciol.* **4**, 173 (1962).
31. Robin, G. de Q. & Adie, R. J. in *Antarctic Research* (eds. Priestley, R., Adie, R. J. & Robin, G. de Q.) 100 (Butterworths, London, 1964).
32. Hughes, T. *Rev. geophys. Space Phys.* **15**, 1 (1977).
33. Doake, C. S. M. *Polar Record* **18**, 37 (1976).
34. Robin, G. de Q. *Norwegian-British-Swedish Antarctic Expedition 1949–52, Scientific Results* **5**, 1–134. (Norsk Polarinstittutt, Oslo, 1958).
35. Tolstikov, Y. I. (ed.) *Atlas Antarktiki*, **1**, 76 (Glavnoye Upravleniye Geodezii i Kartografiy, Moscow and Leningrad, 1966).
36. Nordenskjöld, O. & Mecking, L. *The Geography of the Polar Regions* 72 (American Geographical Society, New York, 1928).
37. Swinbank, C. W. M. *Geograph. J.* **121**, 65 (1955).
38. Bryson, R. A., Wendland, W. M., Ives, J. D. & Andrews, J. T. *Arctic and Alpine Res.* **1**, 1 (1969).
39. Colvill, A. J. *Polar Record* **18**, 390 (1977).
40. Mercer, J. H. *Quat. Res.* **6**, 125 (1976).
41. Emiliani, C. *J. Geol.* **63**, 538 (1955).
42. Shackleton, N. J. & Opdyke, N. D. *Quat. Res.* **3**, 39 (1973).
43. Emiliani, C. *Science* **166**, 1503 (1969). **168**, 1606 (1970).
44. Mercer, J. H. *Science* **168**, 1605 (1970).
45. Voronov, P. S. *Information Bulletin of the Soviet Antarctic Expedition* **23**, 15 (1960) (in Russian).
46. Hollin, J. T. *J. Glaciol.* **4**, 1973 (1962).
47. Mercer, J. H. *Geol. Soc. Am. Bull.* **79**, 471 (1968).
48. Denton, G. H., Borns, H. W., Grosswald, M. G., Stuiver, M. & Nichols, R. L. *Antarctic J. U.S.* **10**, 160 (1973).
49. Weertman, J. *Nature* **260**, 284 (1976).
50. Whillans, I. M. *Nature* **274**, 152 (1976).
51. Bentley, C. R. *Abstracts with Programs* **8**, 773 (Geological Society of America, Denver, 1976).
52. Thomas, R. H. *Nature* **259**, 180 (1976).
53. Weertman, J. *Nature* **260**, 284 (1976).
54. Hughes, T. *Rev. Geophys. Space Phys.* **13**, 502 (1975).
55. Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **194**, 1121 (1976).

Can a myosin molecule bind to two actin filaments?

Gerald Offer & Arthur Elliott

Department of Biophysics, King's College, 26–29 Drury Lane, London WC2, UK

It is suggested that in striated muscles the two heads of one myosin molecule are able to interact with different actin filaments. This would provide a simple explanation for the appearance and arrangement of cross-bridges in insect flight muscle in rigor.

THE myosin molecule has two globular heads attached at one end of a fibrous tail¹. During muscle contraction the heads are thought to bind to actin filaments and produce tension by tilting², but the way in which this is shared by the two heads is not known. It is widely supposed that the two heads of one myosin molecule interact with neighbouring subunits in the

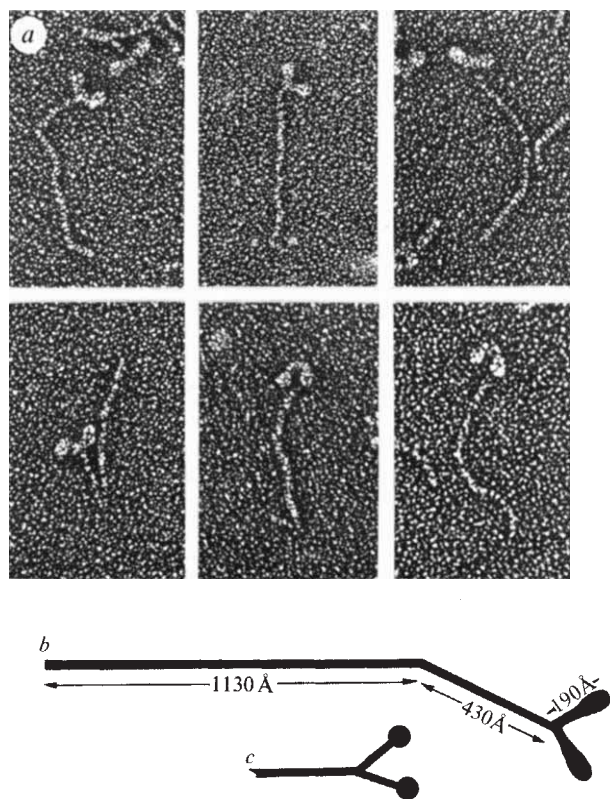


Fig. 1 Shape of the myosin molecule. *a*, Electron micrographs of rabbit skeletal myosin, rotary shadowed with platinum (angle 1 in 5) after drying from 20% ethylene glycol in 0.48 M ammonium formate at -50°C according to the method of Elliott, Offer and Burridge³. Nominal magnification 180,000 platinum coating 14 Å. Full details of the methods and results will be published elsewhere (our manuscript in preparation). *b*, Dimensions of the myosin molecule found from a large number of micrographs like those in (*a*). The diameter of the head is 64 Å at the widest part. *c*, A model in which each head is represented by a ball and spoke of total length 190 Å.

same actin filament. We wish to challenge this view and we show here that because the filament lattice of striated muscle is so well ordered, the two heads of one myosin molecule are capable of interacting with different actin filaments. We were led to make this hypothesis from a redetermination of the shape of the myosin heads. Dr J. Squire and Mr P. Luther have independently reached the same conclusion from a very different approach.

Shape of myosin heads and their attachment to actin

The shape of the myosin molecule is shown in Fig. 1. The heads are 190 Å long, considerably longer than previously thought. They are wider at the end than near the junction with the tail and resemble a pear in shape; for simplicity we have represented each of them in the following discussion as a ball and spoke. The heads are flexibly joined to the tail and can assume a wide variety of tilt angles. Another flexible joint is present in the tail about 430 Å from the head-tail junction as had been postulated earlier by Pepe⁴ and Huxley² for functional reasons.

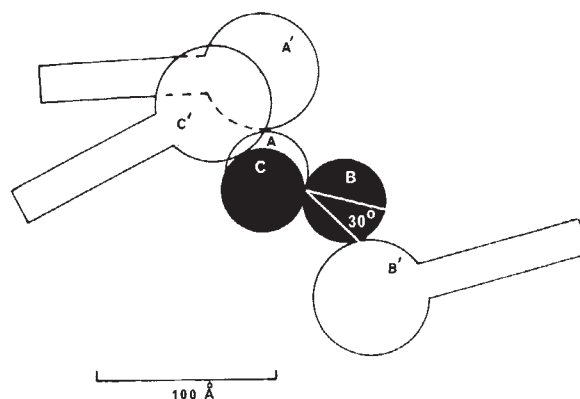
Since heavy meromyosin (the two-headed proteolytic fragment of myosin) binds more strongly to actin filaments than does subfragment 1 (the one-headed fragment)⁵⁻⁷ it is thought that the two heads of a myosin molecule can attach to neighbouring subunits of an actin filament. Consider the geometry of this interaction. The actin filament consists of two long-pitched helical strands of globular subunits^{8,9}. The axial displacement between subunits in the same strand is 55 Å and the pitch is between 700 and 800 Å. Figure 2 shows a transverse slice through an actin filament decorated with myosin heads.

The thickness has been chosen to show one actin subunit in one strand and its neighbours one above and one below in the other strand. The site which each myosin head would occupy if bound in the way described by Moore *et al.*¹⁰ is indicated; we refer to this as a 'head site' without necessarily implying that it is occupied. Because the attached myosin heads follow the same helical symmetry as the actin subunits¹⁰, the distal ends of neighbouring heads are far apart. The distance between two points on a helix of radius r and pitch P separated by an axial distance z is $z[1 + (2\pi r/P)^2]^{1/2}$. For an angle of 50° between the head axis and the thin filament axis, the distal end of the head is at a radius of 160 Å, and the distance between the distal ends of neighbouring heads is therefore 90 Å (72 Å in projection). It follows that when the two heads of one myosin molecule attach to two neighbouring actin subunits they must be distorted in shape so as to meet at the head-tail junction. The deformations will be quite different for the two heads. (That deformations of the heads are possible suggests that part at least of the elasticity present in the cross-bridges¹² may reside in the heads.) This effect would diminish the strength of binding of the second head. That this is indeed the case is supported by recent measurements of the strength of binding of proteolytic fragments of myosin to free F-actin filaments (refs 6, 7 and S. Marston, personal communication). Subfragment 1 binds with an association constant of about 10^7 M^{-1} . The binding of heavy meromyosin is stronger but only by 10- to 100-fold. If the second head could bind with no distortion it would be expected to bind more strongly than the first head because of its enforced proximity to the actin filament; correspondingly the binding constant for heavy meromyosin might have been expected to be at least 10^{14} M^{-1} . Potentially then, the myosin molecule should be capable of binding very tightly indeed to two actin subunits suitably positioned in space, but if the two subunits lie on the same filament, steric constraints greatly reduce the strength of binding.

New model for insect flight muscle in rigor

In contrast to free actin filaments, the actin filaments in striated muscles are arranged in a well-ordered lattice, so interaction of a myosin molecule with actin subunits in two different filaments may be possible. Insect flight muscle is a suitable system for examining this possibility because the filament geometry is better understood than in other muscles, it is better

Fig. 2 Myosin heads attached to one actin filament. The figure shows three neighbouring actin subunits (A, B and C) and the positions A', B' and C' which the attached myosin heads would occupy. The size and shape of the heads are as represented in Fig. 1c. The centre of B and the point of contact of B and B' subtend an angle of 30° at the filament axis¹¹. In projection, we have taken the long axis of the myosin head to be perpendicular to the radius of B' which passes through the point of contact. The long axis of each myosin head is inclined at 50° to the helix axis of the actin filament¹⁰. The overall length of the heads projected on the section plane is 153 Å.



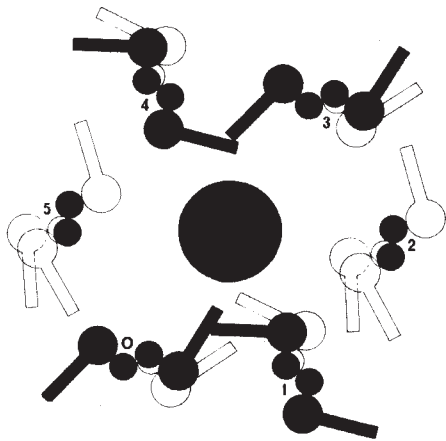


Fig. 3 Pattern of myosin-actin interaction in insect flight muscle in rigor. The diagram shows six actin filaments (depicted as in Fig. 2) surrounding a thick filament (shown by large filled circle). The diameter of the thick filaments (170 Å) and the distance between them (510 Å) are from measurements of electron micrographs and X-ray diffraction^{14,18}. The M line is above and the Z line is below the plane of the diagram.

organised and it operates at an essentially constant interfibrillar distance. We will now consider the structure of this muscle in rigor, the state in which myosin heads are permanently connected to the actin filaments. Reedy and his co-workers¹³⁻¹⁵ have shown that the cross-bridges in this muscle are exceptionally well preserved in electron micrographs and a great deal is known about their shape and arrangement. In thin transverse sections, Y-shaped structures originate at diametrically opposite points on a myosin filament, the pair of structures forming what Reedy has called a "flared cross". The stem of the Y divides about 50 Å from the shaft of the myosin filament and the arms (that is the cross-bridges) are connected to different actin filaments. In his original papers, Reedy¹⁴ pointed out that the Y-shaped structure might contain one, two or four myosin molecules. Subsequently he and other workers have inclined to the view that each Y-shaped structure was formed from two myosin molecules^{16,17}. In the following, we re-examine the possibility that the two heads from a single myosin molecule could form the Y-shape by attaching to different actin filaments. We will refer to this as 'two-filament interaction'. We will then explore the distribution of such interactions in three dimensions.

Figure 3 represents a thin transverse section of insect flight muscle in rigor. In insect flight muscle the actin subunits are arranged on an integral helix with 14 subunits on each helical strand of pitch 770 Å (ref. 18). It is known from Reedy's work that (at least approximately) the decorated actin subunits surrounding a thick filament are arranged in a left-handed helix of pitch 770 Å. This may mean that the actin filaments are themselves arranged in an exact helical manner round the thick filament. This would account for the absence of meridional reflections of spacing 385 and 192 Å (J. Wray, personal communication and ref. 19). An alternative interpretation which would give an approximately helical pattern is that neighbouring actin filaments are related by a rotation of 120°: if as in Fig. 3 actin filaments around a thick filament are numbered 0 to 5 in the anticlockwise direction then 2 is rotated 120° clockwise with respect to 1 and filaments like 1 and 4 which are diametrically opposite are related by a translation in the plane of the figure. We have used the latter interpretation but the conclusions would not be significantly changed if the first had been used.

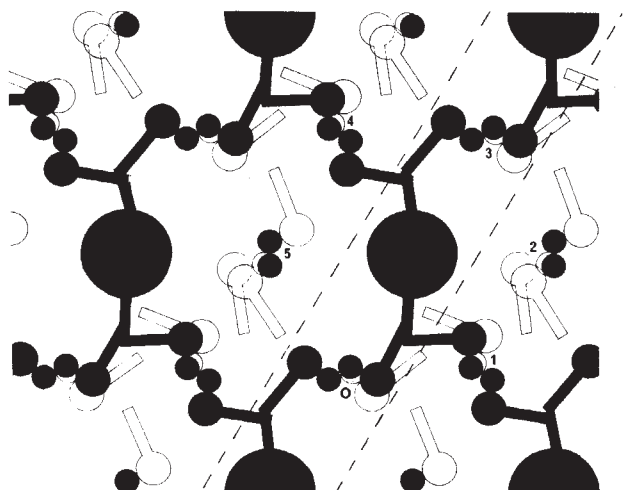
On travelling up the filament axis, the 'head sites' rotate along with the actin subunits, and at certain levels the head sites on neighbouring actin filaments are nearly in contact at their distal ends. Figure 3 shows one such level with these

ends on filaments 0 and 1, as well as on filaments 3 and 4, nearly in contact (they are respectively 35 and 25 Å apart). Such pairs of head sites could therefore be occupied by the heads of one myosin molecule which are shown in solid black. The separations are much less than the 90 Å separation of the distal ends of neighbouring head sites on a single actin filament and the binding would be correspondingly stronger. The head sites on filaments 2 and 5, however, allow no near-contacts with other heads and are not even in a position to allow binding of a single head, as Squire¹⁷ has pointed out.

Those head sites which could be supplied by a single myosin molecule have been shown connected to a single tail in Fig. 4. (The presence of the hinge in the tail would allow the head end of the tail to move out by a radial distance of about 60 Å necessary to allow contact of the heads with the actin filaments.) Thus each attached myosin molecule appears as a Y-shape. Two attached myosin molecules emerging from opposite sides of the thick filament would therefore form a flared cross. Repetition of this structure throughout the lattice generates an appearance strikingly reminiscent of Reedy's micrographs¹⁴. (We assume that the unattached myosin heads are not visible in the electron microscope either because they are not preserved¹⁷ or they lie too close to the shaft to be distinguished from it.) Not only are the shape and size of the cross-bridges much the same but the asymmetry of the cross (resembling the Greek letter χ) predicted by the model can be seen in Reedy's micrographs and has independently been noticed by Reedy and Garrett¹⁵.

The conditions for binding of both heads to separate actin filaments are rather stringent. If the conditions are met at one level, they will not be met for the actin subunits above and below on the same strand, as shown by the unfilled head sites on filaments 0 and 1. We will now consider at what other levels favourable interactions could occur. This is apparently complicated by the fact that the axial separation (145 Å) of myosin along the thick filament^{13,17,18} is different from the axial separation (55 Å) of actin along the thin filament: the whole structure repeats after 2,320 Å (16 myosin translations or 42 actin translations). However, Huxley and Simmons¹² have shown that there is an elastic component in the cross-bridges of vertebrate skeletal muscle which is on average stretched by 60 Å in an isometric tetanus. We might expect therefore that there is sufficient elasticity in the cross-bridges of insect flight muscle to allow axial movements of this magnitude and the hinge in the tail should allow some freedom of azimuthal

Fig. 4 Formation of the 'flared cross' in insect flight muscle in rigor. Similar to Fig. 3 except that a larger area is shown and the myosin heads which could belong to one molecule have been joined to the nearest myosin filament. Each myosin filament is now the centre of an unsymmetrical flared cross whose arms are the cross-bridges. The broken lines denote a thin longitudinal section of muscle which contains a single row of (alternately) actin and myosin filaments (Reedy's 'myac' layer).



position. Initially we shall assume that if any two actin subunits at the same level on different filaments have the right azimuthal relationship, the myosin system will be sufficiently flexible to allow occupation by a myosin molecule.

Thus the problem reduces to determining at what levels neighbouring actin filaments bear the same relationship as filaments 0 and 1, and 3 and 4 in Fig. 3. The right-handed rotation of 360° of the long strands of each actin helix in a distance P (770 Å) together with the rotation of 120° between neighbouring actin filaments result in the pattern of contacts in Fig. 3 repeating approximately at heights $P/6$, $2P/6$, $3P/6$, $4P/6$, $5P/6$ and P after clockwise rotations 60° , 120° , 180° , 240° , 300° and 360° . The relation is approximate because the actin helices have 14 subunits in a length P and so there are not subunits at all of these levels. The nearest occupied levels (at which the pattern appears) are at heights $2P/14$, $5P/14$, $7P/14$, $9P/14$, $12P/14$ and P , which differ from multiples of $P/6$ by only $\pm P/42$ or 0.

In a thin longitudinal section containing one layer of actin and myosin filaments (Reedy's 'myac' layer)¹⁴ the thin filaments will not carry myosin heads at all levels. If the section is taken in the direction shown by the broken line in Fig. 4 (containing actin filaments 0 and 3) there will be two heads on each actin filament near the zero level, corresponding to the plane of the figure. The next level up at $2P/14$ will also be occupied, but at the level $5P/14$ no heads will be attached in this myac layer. The pattern of two occupied levels followed by one unoccupied is shown in Fig. 5 and is similar to Reedy's double chevron pattern^{14,15} if one myosin head is taken to be represented by one bar of a chevron.

In a length of 2,320 Å along the filament axis there are 16 myosin levels but in the surrounding actin filaments there are 18 levels at which attachment could occur. The mismatch between the levels of actin and myosin becomes a maximum twice in this distance, and so an axial period of 1,160 Å may be expected to show either as a missing cross-bridge¹⁵ or by some other feature.

We have so far considered the distribution in space of the two-filament interactions and shown that this satisfactorily accounts for the chief features seen in Reedy's micrographs. These can be explained without postulating other kinds of interactions such as two heads binding to the same actin filament or only one myosin head binding to actin, the other head being unattached. At the levels where two-filament interaction is predicted, we would expect no significant competition from one-filament interactions, for the binding of these would be weaker. Elsewhere, one-filament interaction may occasionally occur but will be restricted by the availability of myosin heads at an appropriate azimuthal¹⁷ and axial position. These must be such that excessive energy is not required to stretch the elastic component and to deform the hinge.

Our model predicts that a smaller fraction of the myosin heads are attached to actin in rigor than has usually been supposed. There are probably six myosin molecules at every level of 145 Å along the thick filament^{16,17,20,21}. If the myosin heads emerging at one level of the thick filament attach at only one level of the actin filament (resulting in one unfilled actin level every 1,160 Å) only 1/3 of the heads would be attached (Fig. 4). If, however, the myosin molecules emerging at one level could reach actin subunits above and below, then the array of double chevrons would be complete and the fraction of myosin heads attached would be slightly greater ($1/3 \times 9/8 = 3/8$).

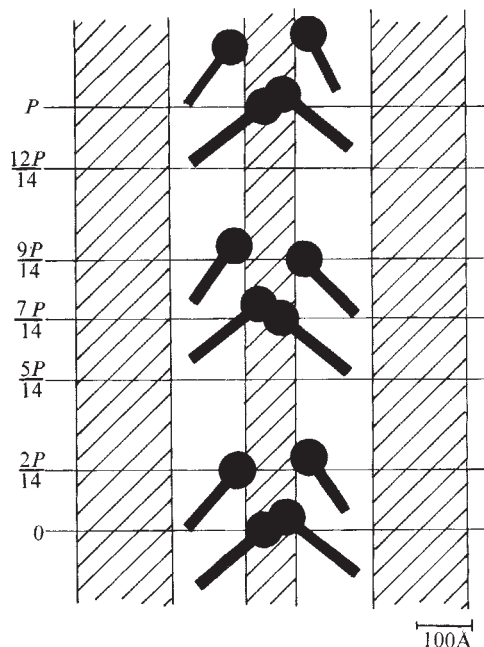
It has usually been assumed that in rigor muscle most, if not all, of the myosin heads are attached^{9,18,22}. Information on the number attached can be obtained from a comparison of the intensities of the two strongest X-ray equatorial reflections, the (1,0) and (2,0), in resting and rigor muscle¹⁸. E. J. O'Brien and J. Couch have calculated that the intensities are consistent with about 11–17% of the mass of the thick filament moving out from close to the shaft and being trans-

ferred to the vicinity of the thin filament when insect flight muscle goes into rigor (personal communication). About half the mass of myosin in a thick filament is in the heads, and the filament contains about 90% myosin²³; hence a movement of 1/3 or 3/8 of the heads corresponds to a mass transfer of 15% or 17%. Both are consistent with the X-ray observations.

Our model accounts for the intensities of the equatorial X-ray reflections in a state of rigor provided that the unattached heads lie close to the thick filament shaft. A satisfactory model should also account for the distribution of intensity among the layer lines, especially the dominance of the 385-Å layer line and the presence of the 128-Å meridional reflection^{18,19,24}. As Reedy¹⁴ has pointed out, the flared cross centred on a myosin filament rotates 60° when the level is increased by about $P/6$ ($P = 770$ Å). The Y-shaped structures around a myosin filament are therefore arranged approximately on a two-stranded helix and have a near-repeat of 385 Å. The 385-Å layer line should therefore be strong and there will be a meridional reflection of spacing $P/6$ (128 Å) (refs 17, 19).

Other models for insect flight muscle in rigor have previously been proposed. The models of Miller and Tregear¹⁸ and of Barrington-Leigh *et al.*²⁴ were devised to account for the intensity of the layer lines in the X-ray diffraction pattern but do not explain the appearance of the cross-bridges in Reedy's electron micrographs. The model of Miller and Tregear (in which, in effect, two-thirds of the actin subunits are decorated with myosin heads) predicts an intensity ratio of the (10) and (20) X-ray equatorial reflections which is very different from the observed. A considerable reduction in the number of heads attached to actin is required to explain the observed intensity ratio. Squire¹⁷ introduced the concept (which we have made use of) that some myosin heads would not be favourably situated for attachment to actin and that these would not be preserved for electron microscopy. He was, therefore, able to predict the appearance of a cross in transverse section. However, his cross-bridges were V-shaped and not Y-shaped since they did not come from a common myosin molecule, and the double chevrons were not accounted for. If both myosin heads from one molecule were bound to adjacent subunits in the same actin filament, each head on Fig. 5 would be accompanied by another separated axially from it by a distance of 55 Å. This would be

Fig. 5 Formation of the 'double chevron' in insect flight muscle in rigor. The thin filament 0 of Fig. 4 is shown in a longitudinal section, with a thick filament on either side. The attached myosin heads are shown filled. The parts of the myosin molecule connecting the heads to the thick filament are not shown.



expected to produce a single very broad chevron and not the clearly separated double chevron. Squire's model, like that of Miller and Tregear¹⁸ predicts that a large fraction of the myosin heads would be attached in rigor and would not be expected to explain the distribution of intensity in the equatorial X-ray diffraction pattern.

In vertebrate skeletal muscle the exact arrangement of neighbouring actin filaments is not known. If they are related by a 90° rotation²⁵, our construction shows that, as in insect flight muscle, the two heads of a myosin molecule may attach to neighbouring actin filaments. In this case, more distortion of the heads would be required to accommodate the varying interfilament distances produced at different sarcomere lengths. Squire and Luther have examined transverse sections through the ends of the A band of vertebrate skeletal muscle in rigor and have concluded that both single-filament and two-filament interactions occur (personal communication).

Conclusions

In biochemical experiments on suspensions of actin filaments the two heads of a myosin molecule probably interact, albeit with strain, with adjacent subunits in the same actin filament. But in the highly ordered filament lattice present in striated muscles considerations of geometry suggest that tighter binding could occur if the two heads bound to different actin filaments.

We have here been concerned with the rigor configuration, which is usually regarded as the termination of the drive-stroke of the cross-bridge cycle², but it would be reasonable to expect that the same two-filament interaction would occur during the earlier part of the stroke. In previous discussions of the mechanism of contraction in muscle, the two-headed nature

of the myosin molecule has largely been ignored. Our ideas suggest a particular way in which the two heads are used; if correct, a new precision will have been given to the hitherto rather vague concept of the cross-bridge.

We are grateful for helpful discussion to Drs M. K. Reedy, R. T. Tregear, J. M. Squire and J. Wray, and to Miss P. M. Bennett, Mr J. Couch and Drs W. B. Gratzer, E. J. O'Brien, J. Trinick and H. White. A.E. acknowledges a project grant from the MRC.

Received 24 June; accepted 14 November 1977.

1. Slayter, H. S. & Lowey, S. *Proc. natn. Acad. Sci. U.S.A.* **58**, 1611–1618 (1967).
2. Huxley, H. E. *Science* **164**, 1356–1366 (1969).
3. Elliott, A., Offer, G. & Burridge, K. *Proc. R. Soc. B* **193**, 43–53 (1976).
4. Pepe, F. A. *J. molec. Biol.* **27**, 227–236 (1967).
5. Young, M. *Proc. natn. Acad. Sci. U.S.A.* **58**, 2393–2400 (1967).
6. Margossian, S. S. & Lowey, S. *J. molec. Biol.* **74**, 313–330 (1973).
7. Highsmith, S. *Biophys. J.* **17**, 40a (1977).
8. Hanson, J. & Lowy, J. *J. molec. Biol.* **6**, 46–60 (1963).
9. Huxley, H. E. & Brown, W. *J. molec. Biol.* **30**, 383–434 (1967).
10. Moore, P. B., Huxley, H. E. & DeRosier, D. J. *J. molec. Biol.* **50**, 279–292 (1970).
11. Parry, D. A. D. & Squire, J. M. *J. molec. Biol.* **75**, 33–55 (1973).
12. Huxley, A. F. & Simmons, R. M. *Nature* **233**, 533–538 (1971).
13. Reedy, M. K., Holmes, K. C. & Tregear, R. T. *Nature* **207**, 1276–1280 (1965).
14. Reedy, M. K. *Amer. Zoologist* **7**, 465–481 (1967); *J. molec. Biol.* **31**, 155–176 (1968).
15. Reedy, M. K. & Garrett, W. E. in *Insect Flight Muscle* (ed. Tregear, R. T.) 115–136 (Elsevier–North Holland, Amsterdam, 1977).
16. Reedy, M. K., Bahr, G. F. & Fischman, D. A. *Cold Spring Harb. Symp. quant. Biol.* **37**, 397–421 (1972).
17. Squire, J. M. *J. molec. Biol.* **72**, 125–138 (1972).
18. Miller, A. & Tregear, R. T. *J. molec. Biol.* **70**, 85–104 (1972).
19. Holmes, K. C., Goody, R. S., Mannherz, H. G., Barrington-Leigh, J. & Rosenbaum, G. in *Mosbach Colloquium on Molecular Basis of Motility* (eds. Heilmeyer, L. M. G., Rüegg, J. C. & Wieland, T.) (Springer, Berlin, 1976).
20. Bullard, B. & Reedy, M. K. *Cold Spring Harb. Symp. quant. Biol.* **37**, 423–428 (1972).
21. Tregear, R. T. & Squire, J. M. *J. molec. Biol.* **77**, 279–290 (1973).
22. Haselgrove, J. & Huxley, H. E. *J. molec. Biol.* **77**, 549–568 (1973).
23. Bullard, B., Luke, B. & Winkelman, L. *J. molec. Biol.* **75**, 359–367 (1973).
24. Barrington-Leigh, J. *et al.* in *Insect Flight Muscle* (ed. Tregear, R. T.) 137–146 (Elsevier–North Holland, Amsterdam, 1977).
25. Squire, J. M. *J. molec. Biol.* **90**, 153–160 (1974).

X-chromosome inactivation during differentiation of female teratocarcinoma stem cells *in vitro*

Gail R. Martin

Department of Anatomy and Cancer Research Institute, University of California, San Francisco, California 94143

Charles J. Epstein, Bruce Travis, Georgianne Tucker & Shaul Yatziv*

Departments of Pediatrics and of Biochemistry and Biophysics, University of California, San Francisco

David W. Martin Jr, Shirley Cliff & Sara Cohen

Departments of Medicine and of Biochemistry and Biophysics, University of California, San Francisco

Evidence is presented that both X chromosomes are genetically active in clonal cultures of undifferentiated female mouse teratocarcinoma stem cells derived from a spontaneous ovarian tumour. As the cells differentiate in vitro one of the X chromosomes becomes inactivated.

EARLY in the development of the female mammalian embryo a process of X-chromosome differentiation, usually termed X-inactivation, occurs in individual embryonic cells^{1,2}. Thereafter, the genes of only one of the two X chromosomes are expressed in that cell and its descendants. As a result dosage compensation occurs so that female cells which have two X chromosomes and male cells which have one produce the same amount of X-linked gene products. While there is ample documentation of this phenomenon, little is known about the mechanism by which it occurs, and there is controversy about the state of activity of the maternally and the paternally derived X chromosomes in female embryos during early post-fertilisation development^{1–3}. The

leading hypothesis is that both X chromosomes are active until some point in development when one or the other is inactivated and ceases to function as a template for messenger RNA (mRNA) synthesis. An alternative view is that the mechanism of X-chromosome differentiation involves not inactivation of an active X chromosome but activation of one of two inactive X chromosomes making the term X inactivation a misnomer.

One of the main obstacles to studies of X-chromosome differentiation is that it occurs very early in embryonic development and appropriate samples for biochemical analysis are consequently difficult to obtain. For example, in the mouse it has been estimated from cytogenetic analyses^{4,5} and cell transfer experiments⁶ that X-chromosome differentiation occurs at approximately the time of implantation (4 d of development) or shortly thereafter when there are relatively few cells in the embryo and it is particularly difficult to remove from the mother. A further difficulty is that male embryos cannot easily be distinguished from female embryos except by cytogenetic analysis, thereby complicating any experiment designed to determine whether one or two X chromosomes are biochemically active in the female embryos of a random population.

*Present address: Department of Pediatrics, Hebrew University, Hadassah Medical Center, Jerusalem, Israel.

Several experimental approaches have indicated that mouse teratocarcinoma stem cells (embryonal carcinoma cells) are very similar to the pluripotent cells of the normal early embryo⁷⁻⁹. It is particularly interesting that certain such cell lines behave like the pluripotent cells of the peri-implantation embryo. Their differentiation *in vitro* is apparently triggered when the cells form rounded aggregates; the first stage in their differentiation, like the first stage in the differentiation of the embryonic inner cell mass, is the formation of a layer of endodermal cells over the free surface of the cell clumps^{10,11}. This suggested that female embryonal carcinoma cells might also be similar to normal embryonic cells with respect to X-chromosome activity and might undergo *in vitro* the changes typical of X-chromosome differentiation. If that were the case, sufficient embryonic material to study this phenomenon could be obtained by using clonal female embryonal carcinoma cells. We report here that a pluripotent female embryonal carcinoma cell line can serve as a model system for the study of X-chromosome differentiation *in vitro*.

Two active X chromosomes in undifferentiated female teratocarcinoma stem cells

We based our experiments on the hypothesis that initially both X chromosomes are genetically active in the embryo and that X-chromosome differentiation is an inactivation process. Therefore, if female teratocarcinoma stem cells are similar to the undifferentiated cells in the embryo at a time before X inacti-

vation they should contain two genetically active X chromosomes. In that case, if both X chromosomes are producing gene products in a dosage dependent manner, the level of any X-linked gene product in female teratocarcinoma stem cells with two X chromosomes should be twice that in comparable male cells with one X chromosome. On the other hand, autosomal gene products should be present in the same amount in female and male cells, because both should contain the same number of copies of these genes.

To obtain appropriate female (XX) and male (XY) stem cell lines we used the methods of Martin and Evans^{10,11} to isolate embryonal carcinoma cells which are maintained in the undifferentiated state by subculture every 3 d in the presence of a confluent feeder layer. These embryonal carcinoma cell lines were isolated from four different tumours: a spontaneous malignant ovarian teratocarcinoma of the LT mouse strain, a spontaneous testicular teratocarcinoma of the Large (Lg) mouse strain and two embryo-derived tumours of the AKR and 129 strains. As expected, the LT cells contained two X chromosomes (Fig. 1). The three other cell lines, however, each contained only one X chromosome but none had a Y chromosome even though one was derived from a spontaneous testicular tumour and was thus definitely male in origin. It should also be noted that none of the cell lines is euploid, although all have a diploid or near-diploid modal chromosome number (Fig. 1). These results are not unexpected for careful cytogenetic analysis usually shows that teratocarcinoma stem cell lines are aneuploid

Table 1 Specific activities of enzymes in undifferentiated teratocarcinoma stem cells

<i>a</i>		LT-XX (clone 1)	PSA1-XO	LG-XO	AKR-XO	Ratio* LT (clone 1) PSA1
X-linked enzymes						
G6PD	Expt 1	148.0	73.0	72.0	75.0	1.93 ± 0.04
	2	52.0	27.0	32.0	25.0	
HGPRT	Expt 1	4.6	2.3	2.4	3.8	2.16 ± 0.07
	2	3.1	1.7	1.6	2.2	
α-gal	Expt 1	5.8	3.2	3.3	3.5	1.73 ± 0.06
	2	5.3	3.2	2.8	2.5	
Autosomal enzymes						
ICDH	Expt 1	20.0	21.0	22.0	20.0	0.94 ± 0.08
	2	46.0	44.0	43.0	40.0	
β-gal	Expt 1	2.5	2.7	2.0	3.2	0.95 ± 0.08
	2	2.8	3.5	2.8	2.5	
β-gluc	Expt 1	2.7	2.9	1.9	1.9	0.86 ± 0.11
	2	1.9	3.0	2.8	2.3	
β-NAC-gluc	Expt 1	75.0	72.0	73.0	58.0	0.99 ± 0.04
	2	68.0	68.0	66.0	70.0	
<i>b</i>		LT-XX (clone 1)	LT-XX (clone 2)	PSA1-XO		Ratio* LT (clone 2) PSA1
X-linked enzymes						
G6PD	Expt 1	162.0	145.0	78.0	1.88 ± 0.02	
	2	192.0	170.0	92.0		
HGPRT	Expt 1	4.0	4.2	2.0	1.94 ± 0.15	
	2	5.8	4.7	2.9		
Autosomal enzymes						
APRT	Expt 1	4.8	3.9	4.6	0.77 ± 0.04	
	2	3.5	3.2	4.5		
AK	Expt 1	1.4	1.5	1.7	0.93 ± 0.11	
	2	1.6	1.6	1.9		
6PGD	Expt 1	95.0	108.0	107.0	1.03 ± 0.04	
	2	133.0	140.0	123.0		

Undifferentiated cells (Fig. 3) were collected by exposure to trypsin at least 3 d after subculture, when in all approximately 5×10^7 cells were collected. Feeder cells represented no more than 10% of the cell population. All cells were washed three times with phosphate-buffered saline and centrifuged. Pellets were used immediately or frozen at -70°C for up to 2 months. At the time of assay cell lysates were prepared by freezing and thawing followed by centrifugation. Specific activities of enzymes were determined as before: G6PD and 6PGD³⁰; HGPRT and APRT³¹; AK (unpublished results of L. Gudas, A. Cohen, B. Ullman and D. W. Martin); ICDH¹⁹; β-gal, β-gluc, β-NAC-gluc³². The assay for α-gal was similar to that for β-gal³², with the appropriate form of the 4-methylumbelliferyl substrate. Each assay was linear with time and protein concentration. *a*, Specific activities of seven enzymes assayed in the cell lines derived from the tumours of all four mouse strains (Fig. 1). The data shown from two independent experiments represent the extreme variation between experiments. *b*, Comparison of specific activities of five enzymes in two independent clones derived from LT tumour 72484. The karyotype of clone 1 is shown in Fig. 1. Clone 2 differs as described in the text.

Specific activities are expressed as nmol product formed per min per mg of protein.

* Ratios given are based on the data from at least three independent experiments, including the two shown. They are reported with their standard errors.

and there is no teratocarcinoma stem cell line in which a Y chromosome has been demonstrated unequivocally¹². In spite of chromosomal abnormalities, the XX-LT cell line was appropriate for our purposes because simple autosomal aneuploidy does not interfere with X-chromosome differentiation². The three XO cell lines were also deemed appropriate because they each contained only one X chromosome and for our purposes could thus be considered 'males'.

To test our hypothesis that both X chromosomes in the LT cells are genetically active, we measured the levels of ten enzymes in the different cell lines. Three of these are known to be X-linked in the mouse^{13,14}: glucose-6-phosphate dehydrogenase (G6PD, EC 1.1.1.49), α -galactosidase (α -gal, EC 3.2.1.22), and hypoxanthine-guanine phosphoribosyl transferase (HGPRT, EC 2.4.2.8). Of the other seven, five are known to be autosomal in the mouse^{14,15}: isocitrate dehydrogenase (ICDH, EC 1.1.1.41), β -galactosidase (β -gal, EC 3.2.1.23), β -glucuronidase (β -gluc, EC 3.2.1.31), adenine phosphoribosyl transferase (APRT, EC 2.4.2.7), 6-phosphogluconate dehydrogenase (6PGD, EC

1.1.1.44); the remaining two are considered to be autosomal by analogy with man: adenosine kinase (AK, EC 2.7.1.20) and β -N-acetyl-glucosaminidase (β -NAC-gluc, EC 3.2.1.30).

The specific activities of these enzymes are shown in Table 1. Although the actual specific activities varied, the ratio of specific activities in the XX-LT cells compared with XO cells was constant for each enzyme from one experiment to the next (Fig. 2). For the six autosomal enzymes tested in all four cell lines, the levels were the same in the XX-LT cells as in the three XO lines. (APRT is not shown because it was not tested in the AKR and Lg cells.) In contrast, the levels of all three X-linked enzymes were twice as high in the XX-LT cells as in the XO cells.

It seems unlikely that these results are related to the aneuploid chromosome constitution of the different cell lines because similar results were obtained when LT cells were compared to three different XO cell lines, each with different chromosomal anomalies (Fig. 1). Moreover, similar results were obtained with a second LT cell line (clone 2, Table 1B) chromosomally different from the first. This cell line, isolated from the same

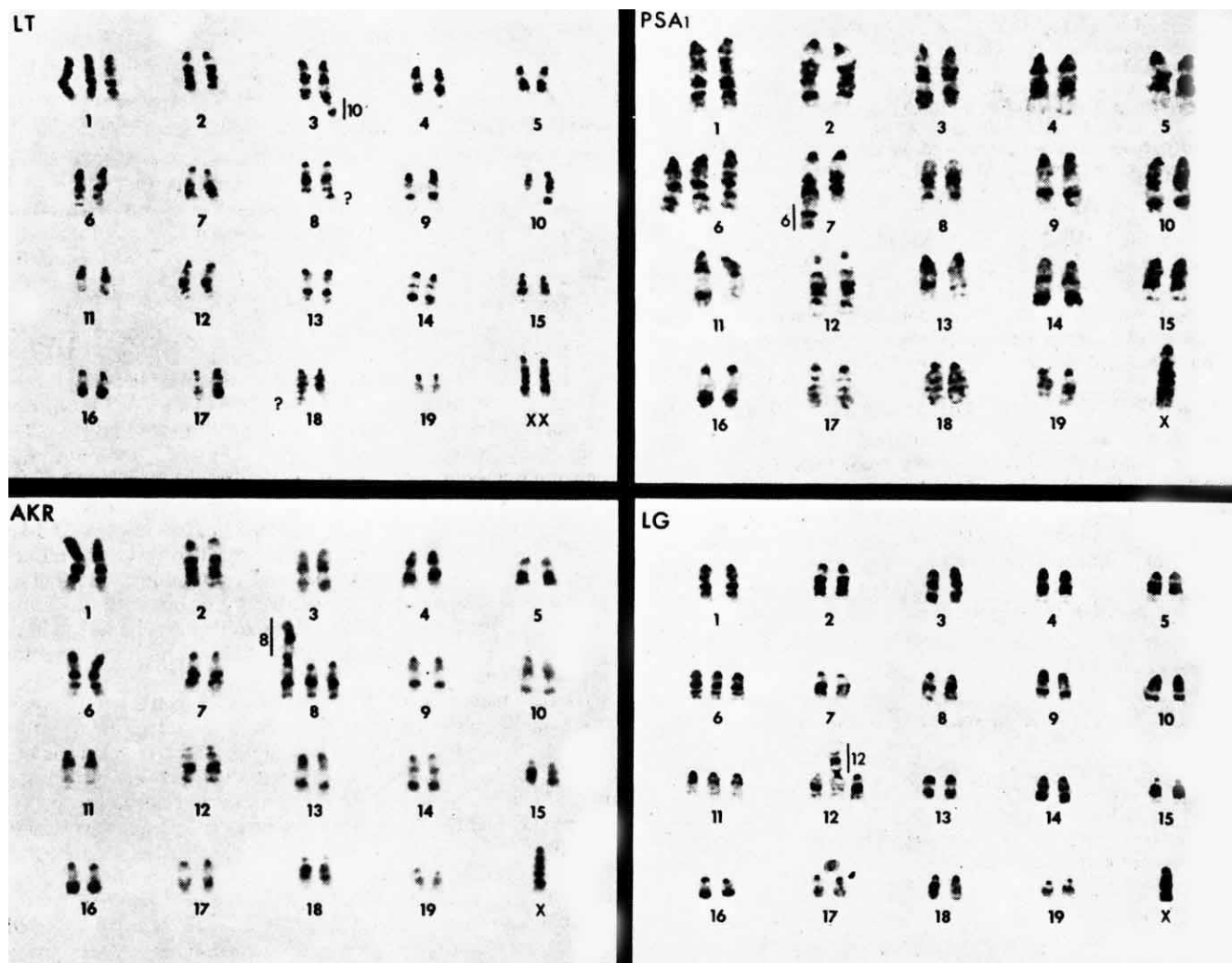


Fig. 1 Karyotypes of teratocarcinoma stem cell lines. Embryonal carcinoma cell lines were isolated as before^{10,11} from the following teratocarcinomas: 34th transplant generation of LT spontaneous ovarian tumour 72482 and 16th transplant generation of Lg strain spontaneous testicular tumour 74218 (both provided by Dr L. C. Stevens, Bar Harbor); 9th transplant generation of AKR strain tumour FA-25-9-1, induced by implantation of a 6.5-d embryo under the kidney capsule of an adult mouse (provided by Dr D. Solter, Wistar Institute); 129 strain tumour OTT5568, induced by implantation of a 3-d embryo in an adult testis²⁷. Each culture derived from these tumours is a homogeneous embryonal carcinoma cell line that can differentiate into various cell types *in vivo* and *in vitro*. Strain 129 (PSA1) cells have been described²⁸. Cells were cultured as described in Fig. 3. Counts of chromosome spreads prepared as described below indicated the modal number of 41 chromosomes in 47/50 LT cells. Modal numbers for the other lines were 40 for PSA1 (50/50 cells examined), 43 for Lg (47/50) and 40 for AKR (48/50). While the banded karyotypes shown here are typical of most cells of each type, there was a significant proportion of the PSA1 cell population in which both number 7 chromosomes were normal. Giemsa-banded karyotypes were prepared as follows: 2–3 d after subculture undifferentiated cells were incubated for 1 h in medium containing 10^{-8} M vinblastine and collected by exposure to trypsin. Cells were suspended in 0.75 M KCl for 20 min and fixed in 1:3 acetic acid-ethanol overnight before being spread on slides. The slides were treated with 6M urea for 20 s at 37 °C and stained in 3% Gurr's Giemsa at pH 6.8 for 5 min. Chromosomes are numbered according to the system of Nesbitt and Francke²⁹.

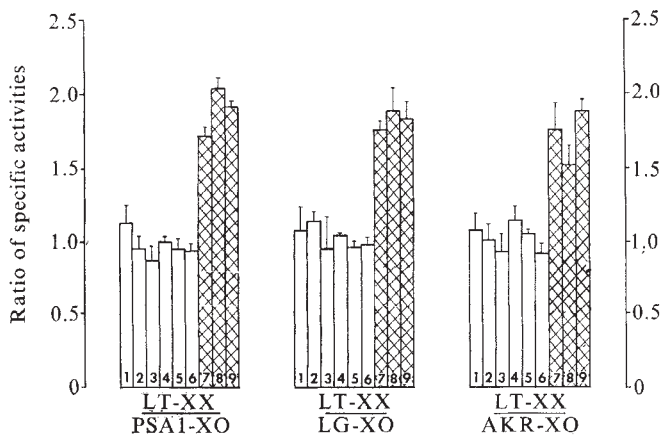


Fig. 2 Ratios of specific activities of autosomal (open columns) and X-linked (hatched columns) enzymes in undifferentiated XX compared with XO teratocarcinoma stem cells. Specific activities of nine enzymes were determined for all four cell lines (Fig. 1). The mean ratio (with s.e.m.) of the specific activity of each enzyme in the LT cells compared with each of the three XO cell lines is shown. Autosomal enzymes: 1, adenosine kinase; 2, β -galactosidase; 3, β -glucuronidase; 4, β -Nac-glucosaminidase; 5, isocitrate dehydrogenase; 6, 6-phosphogluconate dehydrogenase. X-linked enzymes: 7, α -galactosidase; 8, hypoxanthine-guanine phosphoribosyl transferase; 9, glucose-6-phosphate dehydrogenase.

tumour as the LT cells (clone 1) described above, has a modal chromosome number of 40 (trisomic for chromosome 11 and monosomic for chromosome 6, with a few small autosomal rearrangements). These results are consistent with the hypothesis that genes of both X chromosomes are expressed in the undifferentiated female teratocarcinoma stem cells.

X inactivation during differentiation of LT cells

If both X chromosomes are active in undifferentiated LT cells, it can be inferred that these tumour stem cells are similar to embryonic cells at a stage before X-chromosome inactivation. Because the teratocarcinoma stem cells can behave *in vitro* very much like inner cell mass cells during the period when X-chromosome differentiation is probably occurring *in vivo* it seemed likely that these cells would undergo changes in X-chromosome activity if they differentiated *in vitro*. To test this

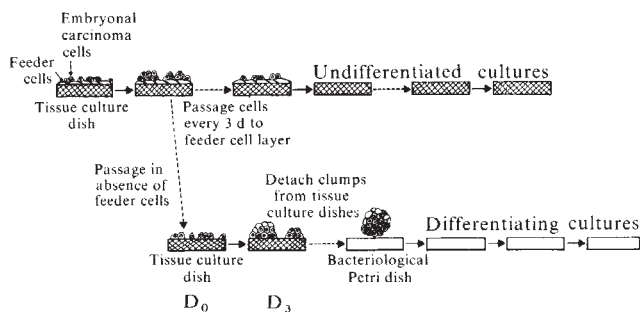


Fig. 3 Culture protocol for teratocarcinoma stem cells. The medium used in all experiments was Dulbecco's modified eagle's medium (containing glucose, 4.5 g l^{-1}) supplemented with 10% calf serum. To maintain embryonal carcinoma cells in the undifferentiated state cells were subcultured every 3 d and seeded at 5×10^6 per 9 cm^2 tissue culture dish containing a confluent layer of mitomycin C-treated STO cells^{10,11}. To initiate differentiation cells were seeded at 10^7 per 9 cm^2 tissue culture dish without feeder cells. After 3 d of culture cell clumps formed and were detached by gentle pipetting, allowed to settle for several minutes, washed in fresh medium and seeded in bacteriological Petri dishes to which they did not adhere. The first differentiated cell type (endoderm) became apparent as a complete outer layer after the clumps were placed in suspension. In some cases differentiation proceeded in the sequence described in ref. 28, but not all the cell lines described here showed exactly that pattern of development.

we cultured both the XX-LT cells and the XO-PSA1 (strain 129) cells in parallel. Separate samples of each were maintained in the undifferentiated state or in conditions conducive to differentiation (Fig. 3). At 2-3-d intervals all cells were assayed and the ratios of specific activities of two X-linked and two autosomal enzymes were determined in the XX compared with XO cells.

The results of a typical experiment (Fig. 4) were those expected if X inactivation occurs during differentiation *in vitro*. A striking change occurred in the ratio of X-linked enzyme activity beginning approximately 9 d after the initiation of differentiation, with the ratio of first G6PD and then of HGPRT dropping from 2.0 to 1.0. As expected, the ratio of activity of autosomal enzymes remained 1.0 whether or not the cells differentiated.

Figure 5 shows that the change in X-linked enzyme activity was specific to the differentiative process. A sample of cells was allowed to differentiate while the parent cultures were maintained in the undifferentiated state. When G6PD activity in the former had decreased, indicating X-chromosome inactivation, a second sample of the undifferentiated parent culture was subcultured in conditions conducive to differentiation. In both cases the decrease in activity began approximately 9 d after the initiation of differentiation. No change in the karyotype of the LT cells occurred during differentiation.

Interpreting the data

The simplest explanation of these data is that both X chromosomes are active in the LT cells and that inactivation occurs when the cells are cultured in conditions conducive to differentiation. This conclusion depends on the validity of the assumption that the amount of gene product present in a cell reflects the number of active genes. More direct measures of X-chromosome activity such as the use of electrophoretic variants of X-linked products are not possible because teratocarcinomas heterozygous for X-linked markers are not yet available. However, there is increasing evidence that, in appropriately controlled conditions, quantification of enzyme activity can provide a valid estimate of the number of functional genes or chromosomes¹⁶⁻¹⁸. The only apparent anomaly in our results involves the autosomal enzyme ICDH, which is governed by the *Id-1* locus on chromosome 1. Recent experiments have shown that trisomy for chromosome 1 results in a proportional increase in ICDH activity in midgestation mouse embryos¹⁹. However, the LT cell line (clone 1) we used is also trisomic for chromosome 1, but no similar dosage effect was noted. There are many possible reasons for the failure to find a dosage effect when one might be expected, and we do not believe this exception invalidates the other findings.

The immediate precedent for our experiments is the assessment of X-chromosome activity in developing and mature oocytes by quantification of gene dosage²⁰⁻²³. In those experiments oocytes from female mice with two X chromosomes had twice the activity of X-linked enzymes as did oocytes from XO females (which had 39 chromosomes, with one X missing). Given that the conclusion, based on gene dosage studies, that both X-chromosomes function in oocytes was confirmed using electrophoretic variants of X-linked markers in human oocytes²⁴, and considering the consistency of our results with all X-linked and control autosomal enzymes except ICDH, it seems likely that our experiments represent a valid approach to the study of X-chromosome function in embryonal carcinoma cells.

In view of the close similarity between the teratocarcinoma stem cells and the pluripotent cells of the normal mouse embryo⁷⁻⁹ these results, which provide the first clear evidence that X-chromosome differentiation is an inactivation process at the molecular level, support the hypothesis that the X-chromosome differentiation during embryonic development involves the biochemical inactivation of an X chromosome.

It is interesting that during differentiation the decrease in activity of G6PD always preceded that of the other X-linked enzyme, HGPRT, by 4 or 5 d. This could be interpreted as an indication that inactivation does not occur along the whole

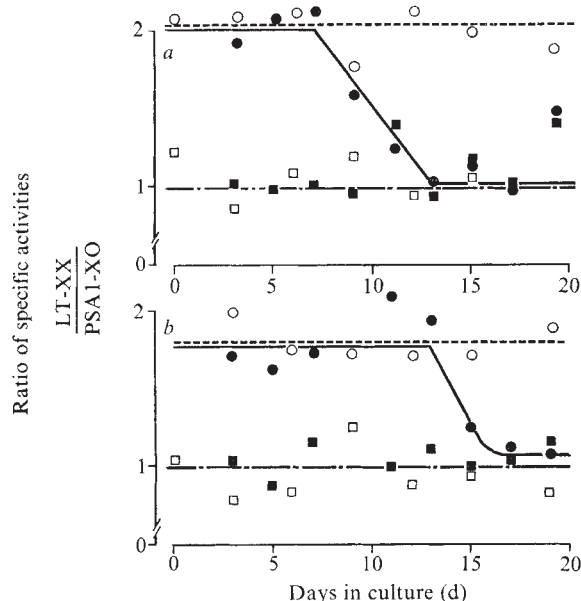


Fig. 4 Levels of X-linked gene products during differentiation of LT cells *in vitro*. Cultures of LT and PSA1 cells were maintained in the undifferentiated state or subcultured (at day 0) in the absence of feeder cells to begin differentiation. At various times during the subsequent 3 weeks, samples of XX-LT and XO-PSA1 cells were collected either by exposure to trypsin followed by centrifugation (for undifferentiated cells) or by centrifugation (for differentiating cultures), and the levels of G6PD, HGPRT, 6PGD and APRT were assayed as described in Table 1. The data shown are typical of the results obtained in two other such experiments. *a*: ○, G6PD in undifferentiated cells; ●, G6PD in differentiating cells; □, 6PGD in undifferentiated cells; ■, 6PGD in differentiating cells. *b*: ○, HGPRT in undifferentiated cells; ●, HGPRT in differentiating cells; □, APRT in undifferentiated cells; ■, APRT in differentiating cells.

length of the X chromosome at one time and that the G6PD locus is closer to an initiation site of X-inactivation than is the HGPRT locus. (The exact location of these genes on the mouse X chromosome is not known.) An alternative explanation is that the whole X chromosome undergoes inactivation at once,

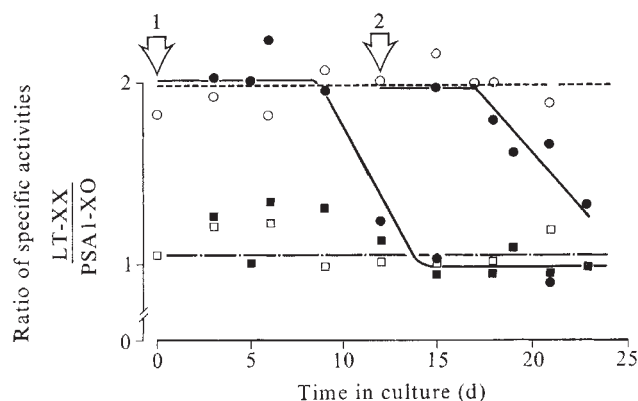


Fig. 5 Relationship between levels of X-linked gene products and the initiation of differentiation. Cultures of LT and PSA1 cells were maintained in the undifferentiated state. At day 0 (arrow 1) one sample of LT-XX cells and one of PSA1-XO cells were subcultured in the absence of feeder cells to initiate differentiation. At day 12 (arrow 2) a second sample each of undifferentiated LT and of PSA1 cells was subcultured in the absence of feeder cells to begin differentiation. Portions of each cell population were collected at various times and the specific activities of G6PD and 6PGD were determined. ○, G6PD in undifferentiated cells; ●, G6PD in differentiating cells; □, 6PGD in undifferentiated cells; ■, 6PGD in differentiating cells.

but that either the mRNA for HGPRT synthesis or the enzyme itself is considerably more stable than that for G6PD. Experiments to distinguish between these possibilities should lead to a better understanding of the molecular mechanism of X-chromosome inactivation.

Our results differ from those of McBurney and Adamson²⁵, who demonstrated by both biochemical and cytogenetic tests that only one of the two X chromosomes appeared to be active in a female teratocarcinoma stem cell line. One possible explanation is that the cells they examined were isolated from a tumour obtained from a 6.5 d embryo and might have been derived from a later (post-inactivation) stage of development than the LT cells described here. A second possibility is that their cell selection and culture procedures, which resulted in feeder-independent stem cells, led to changes in the state of activity of the X chromosomes.

Whatever the reason for this difference, the availability of the LT cells described here makes possible several new approaches to the study of X-chromosome activity. For example, these cells provide a new means of mapping X-linked markers, because the levels of enzymes or even structural proteins suspected of being X-linked can be assayed in the XX-LT and XO-PSA1 cells as described here. LT cells can also be used to study the temporal relationship between biochemical inactivation and the onset of late replication (heterochromatinisation) which is a cytological marker of the inactive X chromosome²⁶. We have determined that in undifferentiated LT cells both X chromosomes replicate at the same time (data not shown). We are investigating whether the changes in enzyme activity found in the LT cells after differentiation are accompanied by changes in the time of replication of one of the two X chromosomes in these cells. The LT cells should also be useful for studies of molecular aspects of X-differentiation, which may provide clues to understanding the way in which genes are regulated during embryonic development.

We thank Dr U. Francke for help with karyotype analysis, Dr G. S. Martin for critical review of the manuscript and Dr Z. Hall for veterinary intensive care. This work was supported by grants from the NSF and the California Division of the American Cancer Society (G.R.M.) and the NIH (D.W.M. and C.J.E.). S.Y. was the recipient of a Fogarty International Fellowship. D.W.M. and C.J.E. are Investigators of the Howard Hughes Medical Institute.

Received 12 September; accepted 17 November 1977.

1. Lyon, M. F. *Biol. Rev.* **47**, 1-35 (1972).
2. Gartler, S. M. & Andina, R. J. *Adv. Human Genet.* **7** (ed. Harris, H. & Hirschhorn, K.) 99-140 (Plenum, New York, 1976).
3. Cattanach, B. M. A. *Rev. Genet.* **9**, 1-18 (1975).
4. Takagi, N. *Expl. Cell Res.* **86**, 127-135 (1974).
5. Mukherjee, A. B. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1608-1611 (1976).
6. Gardner, R. L. & Lyon, M. F. *Nature* **231**, 385-386 (1971).
7. Stevens, L. C. *Adv. Morph.* **6**, 1-31 (1967).
8. Damjanov, I. & Solter, D. *Curr. Top. Path.* **59**, 69-130 (1974).
9. Martin, G. R. *Cell* **5**, 229-243 (1975).
10. Martin, G. R. & Evans, M. J. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1441-1445 (1975).
11. Martin, G. R. & Evans, M. J. *Roche Symposium on Teratomas and Differentiation* (ed. Sherman, M. I. & Solter, D.) 169-187 (Academic, New York, 1975).
12. Graham, C. F. in *Concepts in Mammalian Embryogenesis* (ed. Sherman, M.) 315-394 (MIT Press, Cambridge, 1977).
13. Chapman, V. M. & Shows, T. B. *Nature* **259**, 665-667 (1976).
14. Kozak, C., Nichols, E. & Ruddle, F. H. *Somatic Cell Genet.* **1**, 371-383 (1975).
15. *Mouse News Letter* **56**, 4-28 (1977).
16. Marimo, B. & Giannelli, F. *Nature* **256**, 204-206 (1975).
17. Magenis, R. E., Koler, R. D., Lovrien, E., Bigley, R. H., DuVal, M. C. & Overton, K. M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4526-4530 (1975).
18. George, D. L. & Francke, U. *Science* **194**, 851-852 (1976).
19. Epstein, C. J., Tucker, G., Travis, B. & Gropp, A. *Nature* **267**, 615-616 (1977).
20. Epstein, C. J. *Science* **163**, 1078-1079 (1969).
21. Epstein, C. J. *Science* **175**, 1467-1468 (1972).
22. Maugia, V., Abbo-Halbasch, G. & Epstein, C. J. *Dev. Biol.* **45**, 366-368 (1975).
23. Kozak, L. P., McLean, G. M. & Eichler, E. M. *Biochem. Genet.* **11**, 41-47 (1974).
24. Gartler, S. M., Liskay, R. M., Campbell, B. K., Sparkes, R. & Gant, N. *Cell Diff.* **1**, 215-218 (1972).
25. McBurney, M. W. & Adamson, E. D. *Cell* **9**, 57-70 (1976).
26. Ohno, S. & Cattanach, B. M. *Cytogenetics* **1**, 129-140 (1962).
27. Stevens, L. C. *Dev. Biol.* **21**, 364-382 (1970).
28. Martin, G. R., Wiley, L. M. & Damjanov, I. *Dev. Biol.* **61**, 69-83 (1977).
29. Nesbitt, M. N. & Francke, U. *Chromosoma* **41**, 145-158 (1973).
30. Glock, G. E. & McLean, P. *Biochem. J.* **55**, 400-408 (1953).
31. Green, C. & Martin, D. W., Jr. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3698-3702 (1973).
32. Sandman, R., Margules, R. M. & Kountz, S. L. *Clin. Chim. Acta* **45**, 349-359 (1973).

letters to nature

Is Fairall 9 an X-ray Seyfert galaxy?

FAIRALL recently identified a new Seyfert galaxy, designated F-9, from coarse spectra of a galaxy evident on an ESO Fast Blue Survey Plate¹. Only limited data were presented, however, for example, a redshift was not given, nor was the galaxy spectroscopically classified (Seyfert type 1 or 2)². In a search at MIT for possible catalogued optical counterparts to X-ray sources, I noticed that F-9 lies within the common intersection of both a 2A and 4U X-ray error box^{3, 4}. Figure 1 shows the 2A error box and part of the (much larger) 4U error box overlaid on an ESO Fast Blue Survey Plate of the region around F-9. The apparent magnitude of F-9 is given as mag ~ 13 (ref. 1). Because of its fuzzy perimeter, it is the most conspicuous object in the 2A error box.

All known X-ray emitting Seyfert galaxies are of type 1 (refs 5 and 6). A determination of the optical classification (type 1 or 2), as well as a measurement of the redshift of F-9 is desirable. Such information, combined with an improved X-ray error box, is essential to establishing the nature of F-9.

After the submission of this paper, West⁷ and Ward *et al.*⁸ reported a determination of the spectroscopic class of Fairall-9 (= ESO113-IG45). They find it to be a type 1 Seyfert, with a redshift $z = 0.045$ and magnitude $V = 13.2$. Thus, F-9 seems to be the brightest Seyfert galaxy known in the Southern Hemisphere. If the identification of F-9 with 2AO120-591 (4U0105-59) suggested here is correct, then its X-ray luminosity (2-10 keV) is approximately 3.10^{44} erg s⁻¹, which is com-

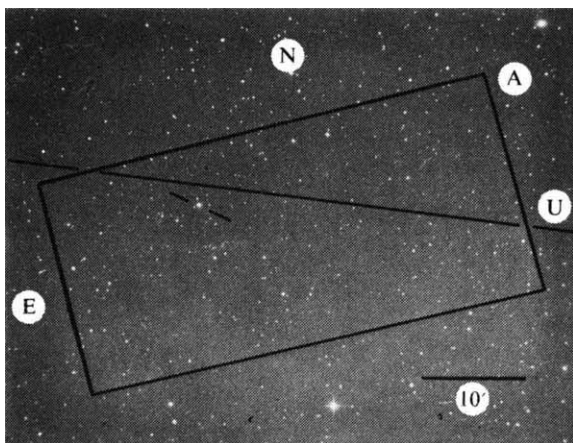


Fig. 1 The X-ray source 2AO120-591 (4U0105-59) and the Seyfert galaxy F-9. 'A' designates the 2A error box³ and 'U' the northern edge of the 4U error box⁴. The nucleus of F-9 is at α (1950) = 01 h 21.8 min, δ (1950) = $-59^{\circ}04'$ (ref. 1). The photograph is reproduced from an ESO Fast Blue Survey Plate.

parable to that of the brighter X-ray emitting Seyfert galaxies^{5, 6}. Detailed studies of this unique object at radio, IR, optical and X-ray wavelengths are obviously needed, particularly with a view to establishing variability.

This work was supported in part by NASA under contract NAS5-11450 and the USNSF under Grant MPS 75-02963.

G. R. RICKER

*Department of Physics and Center for Space Research,
Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139*

Received 12 October; accepted 28 November 1977.

1. Fairall, A. P. *Mon. Not. R. astr. Soc.* **180**, 391 (1977).
2. Weedman, D. W. A. *Rev. Astr. Astrophys.* **15** (in the press).
3. Cooke, B. A. *et al.* *Mon. Not. R. astr. Soc.* (in the press).
4. Format, W. *et al.* *Astrophys. J. Suppl.* (in the press).
5. Elvis, M. *et al.* *Mon. Not. R. astr. Soc.* (in the press).
6. Tananbaum, H. *et al.* *Astrophys. J. Lett.* (in the press).
7. West, R. M. *IAU Circ.* No. 3134, 3143, 1977.
8. Ward, M. J. *et al.* preprint (to be submitted to *Mon. Not. R. astr. Soc.*).

Infrared light curves of AM Herculis

AM HER has been identified as the optical counterpart of 3U 1809+50 by its common period of 3.1 h (refs 1, 2). The system shows strong emission lines, the radial velocities of which have been measured by Priedhorsky³ and Cowley⁴. Tapia² has measured strong linear and circular polarisation which indicates that one component of the system is probably a white dwarf emitting strong cyclotron radiation. Photometric light curves have been observed^{5, 6} which show two minima, unlike the X-ray light curve which has only one; the system also has a red colour index. Models of the system which rely on a white dwarf, accretion disk, hot spot, and a gas stream from a companion star that fills its Roche Lobe have been proposed by Channugan⁷ and Fabian⁸. In view of the red colour, strong emission lines and cyclotron radiation, we considered that infrared observations of this star would be of interest.

The 2.2 μ m light curve was observed on the night of the 9-10 July 1977 and the 1.2 μ m curve was observed the following night. The observations were made with the 1.5 m Infrared Flux Collector at the Cabezon Observatory, Tenerife. A standard infrared photometer with an InSb detector operated at 63 K was used. Observations covered one period only on both nights but there is a short break in the 2.2 μ m light curve when it was thought the cryostat had run out of liquid nitrogen. Both nights were of excellent photometric quality and α Lyr, used as the standard star, was monitored throughout the night, but only before and after the AM Her observations. This procedure was followed so that the light curves should be as complete as possible. The magnitudes assumed for α Lyr are $J = 0.02$ and $K = 0.02$ (ref. 9). Corrections were made for atmospheric extinction but these were very small. Each observation is the result of a 60-s integration, 30 s in each channel. The errors on each point, estimated from observations of a similarly faint non-variable star, are ± 0.05 mag. Thus the spread of the points is thought to be genuine flickering of the star. Such flickering is also seen in the optical light curves^{5, 6}.

The light curves are shown in Fig. 1. The times shown are heliocentric Julian dates. Also shown are the positions of primary and secondary minima expected from the photometric ephemeris⁶ H.J.D. $2,443,014.7133 + 0.128924 \times E$.

It is unfortunate that we did not observe the bottom of the 2.2 μm secondary eclipse. However, the whole interest of these observations would seem to be the absence of the 2.2 μm primary eclipse. Thus the 2.2 μm light curve is similar to the X-ray light curve which also shows only a secondary minimum. In fact, there may be a shallow 2.2 μm primary eclipse, hidden by the flickering; if the data is smoothed a depth of 0.15 mag would be consistent with the results.

The 1.2 μm light curve is very similar to the optical light curve and so could be explained by any model which describes the optical light curve. The model of Fabian *et al.*⁸ explains

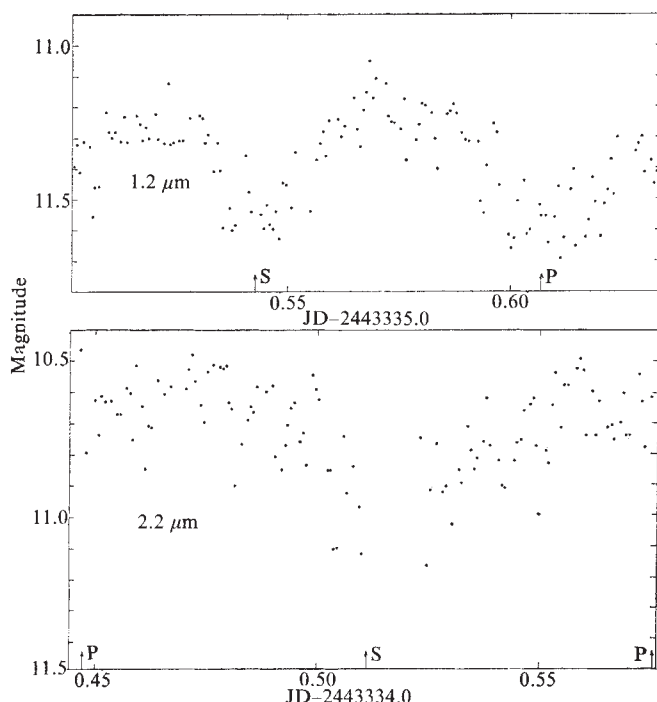


Fig. 1 1.2 μm and 2.2 μm light curves of AM Herculis

primary minimum as the eclipse of the hot spot by the companion star. This hot spot is formed where the gas stream meets the accretion disk; it is also a high density region. The secondary minimum is the eclipse of the white dwarf by this high density hot spot. X-rays are formed by accretion of material on to the white dwarf. The temperature of the hot spot is not sufficient to produce X-rays and thus the X-ray light curve shows only the secondary minimum. The similarity of the 2.2 μm and X-ray light curves suggest that the dominant 2.2 μm emission must also arise near the white dwarf stars.

We note, also, that the system has a positive colour index, $J-K \simeq 0.60$ at maximum light. This could be produced by the cool companion or free-free radiation from the accretion disk, however, it is difficult to see how either of these regions could be suitably eclipsed. It is known that the white dwarf emits cyclotron radiation generated by its strong magnetic field.

With cyclotron radiation, different frequencies are emitted at different magnetic fields and so in different volumes. This can lead to a flatter spectrum than black body emission, which might produce the observed colour index. We therefore propose that 2.2 μm light curve is explained by strong cyclotron radiation produced near the white dwarf.

We are grateful to the SRC for travel funds to Tenerife. A.M. has an SRC studentship. We also thank Drs A. R. King and D. J. Raine for helpful discussions.

R. F. JAMESON
R. AKINCI
D. J. ADAMS
A. B. GILES

Department of Astronomy,
Leicester University, UK

A. MCCALL

Department of Astronomy,
Hatfield Polytechnic, UK

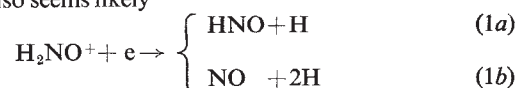
Received 15 August; accepted 23 November 1977.

- Hearn, D. R. & Richardson, J. A. *MIT preprint* CSR-P-76-36, (1976).
- Tapia, S. *Astrophys. J. Lett.* **212**, L125 (1977).
- Priedhorsky, W. C. *Astrophys. J. Lett.* **212**, L117 (1976).
- Cowley, A. P. & Crampton, D. *Astrophys. J. Lett.* **212**, L121 (1976).
- Szkody, P. & Brownlee, D. E. *Astrophys. J. Lett.* **212**, L113 (1976).
- Olson, E. C. *Astrophys. J.* **215**, 166 (1976).
- Channugan, G. & Wagner, R. L. *Astrophys. J. Lett.* **213**, L13 (1977).
- Fabian, A. C., Pringle, J. E., Rees, M. J. & Whelan, J. A. *J. Mon. Not. R. astr. Soc.* **179**, 9P (1977).
- Johnson, H. L., Mitchell, R. I., Iriarte, B. & Wisniewski, W. Z. *Commun. lunar planet. Lab. Univ. Arizona* **4**, 99 (1966).

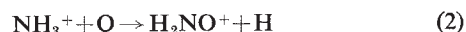
Interstellar nitroxyl

ULICH, Hollis and Snyder¹ have reported detection of the molecule nitroxyl, HNO, in the interstellar clouds SgrB2 and NGC2024. The detection is important as an addition to the long list of molecules known to be present in interstellar clouds, but also more particularly because it is the only interstellar molecule so far detected which contains the NO bond. The failure to detect this bond has been an important observational fact with which theoretical formation schemes have had to conform. Ulich *et al.*¹ also briefly discussed formation processes for HNO. We² have considered a large network of reactions involving the chemistry of interstellar nitrogen, and we are able to estimate the densities of some species which take part in reactions affecting HNO formation and loss and which are reported here.

HNO formation may arise from the recombination (1a), though (1b) also seems likely³



where the H_2NO^+ itself is formed from



(A. Dalgarno, see ref. 1; also ref. 3). NH_3^+ arises in several ways; Watson⁴ has suggested starting reactions involving either

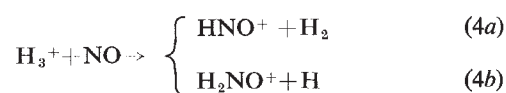


or

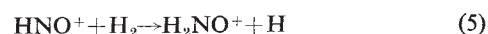


where these ions eventually end up as NH_3^+ . In addition, grains may supply NH_3 directly to the gas; charge exchange^{5,6} between C^+ and NH_3 is another source for NH_3^+ , though this seems to be of minor importance in cloud interiors².

It seems worth examining other possibilities, for example,



and (4a) would need to be followed quickly by



The reaction of NO^+ and H_2 to give HNO^+ apparently does not proceed⁷.

Reactions on grain surfaces may contribute directly to HNO formation. Allen and Robinson⁸ give estimates of HNO densities in various clouds as a function of time. The model of Pickles and Williams² also enables calculation of the HNO formation rate. Loss of HNO seems likely with ions such as HCO^+ , C^+ , H^+ , He^+ , and H_3^+ , and also in reaction with neutrals such as O- and C- atoms. Loss by collision with sticking to grains may also occur.

We present in Table 1 formation and destruction rates for HNO based on densities of reactants calculated in our 'dense cloud' model², in which $n(\text{H}_2) = 10^4 \text{ cm}^{-3}$. This may be the appropriate density for both Sgr B2 (see ref. 9) and NGC 2024 (see ref. 10) unless severe clumping is occurring with HNO. Formation by equation (1a) is assumed to proceed at the Langevin rate, and (1b) is suppressed. It has been shown², in any case, that (1b) is unlikely to lead efficiently to interstellar NO formation. The calculated NH_3^+ densities lie in the range 7×10^{-6} to $5 \times 10^{-5} \text{ cm}^{-3}$ and typically $n(\text{O}) \sim 0.4 \text{ cm}^{-3}$. Contributions to NH_3^+ from (3a) and (3b) are included. The main determinators for NH_3^+ in cloud interiors are formed from NH_2^+ and loss by reaction with atomic oxygen and by dissociative recombination. Also we note that the scheme implies an H_2NO^+ abundance as high as $3 \times 10^{-6} \text{ cm}^{-3}$, if we assume loss occurs by recombination at the usual rate.

Table 1 Formation and destruction rates of HNO

Formation

Initiating reaction

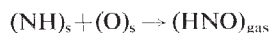
$\text{NH}_3^+ + \text{O}$	$3 \times 10^{-15} - 2 \times 10^{-14} \text{ cm}^{-3} \text{ s}^{-1}$
$\text{H}_3^+ + \text{NO}$	$\lesssim 10^{-16} \text{ cm}^{-3} \text{ s}^{-1}$
$\text{NO}^+ + \text{H}_2$	0
Grains	$3 \times 10^{-15} \text{ cm}^{-3} \text{ s}^{-1}$

Destruction

Route	
Reaction with ions	$10^{-14} - 2 \times 10^{-12} \text{ s}^{-1}$
Reaction with neutral O atoms	$4 \times 10^{-12} \text{ s}^{-1}$
Sticking to grains	$2 \times 10^{-13} \text{ s}^{-1}$

Huntress⁷ argues that equation (4a) is the more likely channel in the $\text{H}_3^+ + \text{NO}$ reaction, if it is exothermic. Reaction (5) must then be fast. The proton affinities of Huntress⁷ indicate that reaction (4a) is exothermic (but the opposite is true, if we follow ref. 11). Assuming that equation (4a) proceeds at the Langevin rate, is followed rapidly by reactions (5) and (1a), and using the observational upper limit for NO, it gives the value in Table 1.

Formation on grains may occur in such reactions between surface species as



which are exothermic⁸. The value quoted is from our data but Allen and Robinson⁸ quote a density which implies a formation rate several times larger than this.

Loss of HNO: most models^{2,12} have ion densities $10^{-4} - 10^{-5} \text{ cm}^{-3}$ for $n \sim 10^4 \text{ cm}^{-3}$. Reactions with neutral atomic oxygen with $n(\text{O}) \sim 0.4 \text{ cm}^{-3}$, are assumed to occur at the collisional rate. These loss rates are sufficiently fast for equilibrium to be reached on a time scale which is much less than the free fall time for the density assumed here.

The observations imply $\text{N}(\text{HNO})/\text{N}(\text{H} + 2\text{H}_2) \sim 10^{-9}$ in both sources. Ignoring clumping, and assuming HNO is uniform with hydrogen, then $n(\text{HNO}) \sim 10^{-5} \text{ cm}^{-3}$. The rates in Table 1 seem to indicate a density greater than this by at least two orders of magnitude. As the loss rates are unlikely to be faster, we infer that all the suggested theories overestimate the formation rates. We require that both gas phase reaction (1a) and the grain surface reaction be much less efficient than assumed here, as both seem to contribute equally. The abundance

calculated here is fairly insensitive to $n(\text{O})$ because it appears in formation and loss.

We conclude first that both gas and grain formation routes may be important for HNO, as is apparently the case for other molecules², and second that theories seem massively to overestimate the abundance of some molecules in high density clouds. Both these conclusions were also reached in considering chemistry of other molecules². The apparent overproduction of certain molecules in high density clouds by both gas and grain mechanisms does not generally seem to be appreciated.

J. B. PICKLES
D. A. WILLIAMS

Mathematics Department,
University of Manchester,
Institute of Science and Technology,
Manchester, UK

Received 14 November; accepted 5 December 1977.

1. Ulich, B. L., Hollis, J. M. & Snyder, L. E. *Astrophys. J. Lett.* **217**, L105 (1977).
2. Pickles, J. B. & Williams, D. A. *Astrophys. Space Sci.* (in the press).
3. Barsuhn, J. *Astr. Astrophys. Suppl.* **28**, 453 (1977).
4. Watson, W. D. *Astrophys. J. Lett.* **183**, L17 (1973).
5. Huntress, W. T. & Anicich, V. G. *Astrophys. J.* **208**, 237 (1976).
6. Schiff, H. I., Hemsworth, R. S., Payzant, J. D. & Bohme, D. K. *Astrophys. J. Lett.* **191**, L49 (1974).
7. Huntress, W. T. *Astrophys. J. Suppl.* **33**, 495 (1977).
8. Allen, M. & Robinson, G. W. *Astrophys. J.* **212**, 396 (1977).
9. Scoville, N. Z., Solomon, P. M. & Penzias, A. A. *Astrophys. J.* **201**, 352 (1977).
10. Grasdalén, G. L. *Astrophys. J.* **193**, 373 (1974).
11. Watson, W. D. in *Atomic and Molecular Physics and the Interstellar Matter* (eds Balian, R., Encrenaz, P. & Lequeux, J.) (Elsevier, North-Holland, Amsterdam, 1974).
12. Herbst, E. & Klemperer, W. *Astrophys. J.* **185**, 505 (1973).

Sorption of phosphate by hydrolytic reaction products of aluminium

THE practical importance of hydrated aluminium oxides in the terrestrial and aquatic environments is well known primarily through the significant role they have in the retention of nutrients in agriculture¹ and in the control of pollutants such as phosphate in eutrophication processes². Anion adsorption on hydrous aluminium oxides has also attracted the interest of engineers mainly because in tertiary waste-water treatment, the removal of phosphate with alum salts seems to be through adsorption on the hydrated oxide floc particles^{3,4}. Consequently the adsorption of anions on hydrated aluminium oxides has been extensively investigated. These studies have shown that organic anions such as citrate compete with phosphate for adsorption sites on the surfaces of the hydrated aluminium oxides thereby decreasing their capacity to retain phosphate⁵⁻⁸. Our studies⁹ show that citric acid even at very low concentrations hampers the formation of crystalline aluminium hydroxides. We report here that due to the hindrance of the crystallisation processes, the retention of phosphate by the hydrolytic reaction products of aluminium is enhanced and not reduced when they are precipitated in the presence of low concentration of citric acid which represents an organic acid commonly occurring in nature.

Table 1 shows the retention of phosphate by the hydrolytic precipitation products of aluminium increased with concentration of citric acid present during their formation. After 40 d, for instance, the solid phase products formed in the presence of 10^{-6} M citric acid retained, over the pH range 4-8, about 200% more phosphate than would have precipitated in the absence of citric acid. Furthermore when the citric acid concentration was raised to 10^{-4} M , the phosphate retained by the resulting hydrolytic precipitation products was increased by as much as 300-400% over the pH range investigated. The high retention capacities of phosphate by the hydrolytic reaction products of aluminium, precipitated in the presence of citric acid, are attributed to their noncrystalline nature. Citric acid chelates

Table 1 Retention of phosphate by hydrolytic precipitation products of aluminium collected from systems aged for 40 d in the absence and in the presence of citric acid

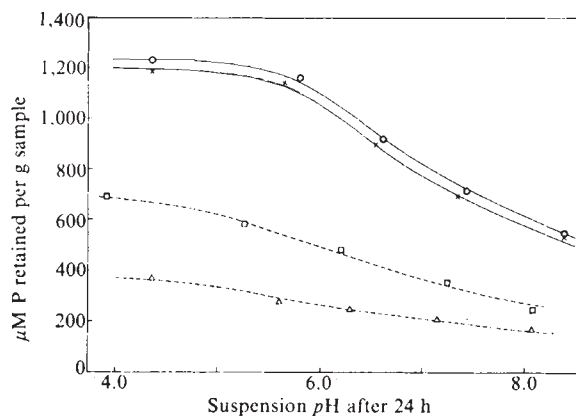
Suspension pH at equilibrium	$\mu\text{M P}$ retained per g sample		
	Citric acid concentration (M)		
	0	10^{-6}	10^{-4}
4.0	370	745	1,205
5.0	330	685	1,180
6.0	270	560	1,060
7.0	210	435	775
8.0	165	335	575

The preparation of the hydrolytic precipitation products of aluminium starting from an initial aluminium concentration of 1.10×10^{-3} M and the OH/Al molar ratio of 3.0 in the absence and in the presence of 10^{-6} M and 10^{-4} M citric acid has been described elsewhere⁹. Retention of phosphate was determined by shaking 50 mg of precipitation products in 50 ml of 0.1 M KCl solution containing 1.0×10^{-5} M KH_2PO_4 for 24 h at 25 °C. The data are deduced from a plot of $\mu\text{M P}$ retained per g sample against pH of the equilibrium suspension.

aluminium ions in aqueous solution^{10,11} and as a result of the occupation of coordination sites of aluminium by citrate during the formation of the precipitated products, distortion in the arrangement of hexagonal ring units normally found in crystalline aluminium hydroxides is bound to occur. The distortion as shown by X-ray diffraction analyses and electron microscopic observations was sufficient to impart a rough surface to the precipitation products and to cause the products formed in the presence of 10^{-6} – 10^{-4} M citric acid to be noncrystalline to X rays⁹. The mechanisms of adsorption of phosphate on hydrated aluminium oxides, on the basis of existing information¹², involve either the displacement of H_2O at positive sites or OH at negative sites. The formation of these phosphate retention sites on the surfaces of hydrolytic reaction products of aluminium would thus be promoted by the structural distortion effect of citric acid during the hydrolytic precipitation process. Moreover, the noncrystalline nature of the products enhances their specific surface (Table 2). This increase in specific surface of the products coupled with their noncrystalline nature and thus reactive sites for phosphate retention would apparently, at the low concentrations of citric acid, more than compensate the blocking of retention sites by citric acid and explains why the retention of phosphate by the products precipitated in the presence of citric acid is more extensive than in its absence.

The capacity of the noncrystalline solid phase products precipitated in the presence of 10^{-4} M citric acid to retain phosphate is not substantially diminished by an ageing period of 40 d (Fig. 1). On the other hand, the uptake of phosphate by the corresponding products formed in the absence of citric acid is considerably reduced after 40 d, for example, from 675 to 370 $\mu\text{M P}$ per g at pH 4.0 or from 270 to 165 $\mu\text{M P}$ per g at pH 8.0 (Fig. 1). Noncrystalline

aluminium hydroxides precipitated in the absence of interfering ions are very unstable and revert rapidly to the crystalline state¹⁴. On reversion to their crystalline state the specific surface of the aluminium hydroxides is considerably reduced (Table 2) with a concomitant decrease in the active sites for phosphate retention. In the presence of citric acid, the occupation of coordination sites by citrate would retard the reversion of the precipitation products to the crystalline state as that would require the removal of citrate from within the structure. The maintenance of the noncrystalline state with an accompanying high specific surface (Table 2) by citric acid would thus help to stabilise the high phosphate retention capacities of noncrystalline aluminium hydroxides.

**Fig. 1** Retention of phosphate by hydrolytic reaction products of aluminium formed in systems at the initial aluminium concentration of 1.10×10^{-3} M and OH/Al molar ratio of 3.0 in the absence (dashed line) and in the presence of 10^{-4} M citric acid (solid line) as a function of pH and time. Phosphate retention was determined as outlined in Table 1. ○ and □, after 1 d; × and △, after 40 d.

Under natural environments, absolute elimination of low molecular weight biochemicals, such as citric acid, is seldom achieved and it is more realistic to believe that the hydrolytic reaction products of aluminium in soils and aquatic sediments are generally formed in the presence of minute but measurable quantities of biochemical compounds. Previous investigations⁵⁻⁸ have only reported the tendency of these biochemical compounds to reduce anion uptake by hydrated aluminium oxides through competition with the anions for adsorption sites. The present study provides direct evidence that they also enhance the retention of anions such as phosphate by the hydrolytic reaction products of aluminium through their promoting effects on the formation of noncrystalline aluminium hydroxides. In nature, therefore, the low molecular weight organic compounds have a dual role of hindering and promoting anion retention by the hydrated aluminium oxides depending on the mechanisms of their interfering reactions. This finding should be of fundamental importance not only in understanding the basic colloidal chemistry of aluminium but also in the transport and fate of nutrient and pollutant anions in terrestrial and aquatic ecosystems.

This work was supported by a grant from the NRC of Canada.

K. F. NG KEE KWONG
P. M. HUANG

Saskatchewan Institute of Pedology,
University of Saskatchewan,
Saskatoon, Saskatchewan, Canada S7N 0W0

Table 2 Specific surface of the hydrolytic precipitation products of aluminium in the absence and in the presence of citric acid, after 1 d and 40 d

Initial citric acid concentration (M)	Specific surface ($\text{m}^2 \text{g}^{-1}$)	
	Ageing period (d)	
	1	40
0	109.0	18.9
10^{-6}	136.6	117.1
10^{-4}	311.5	295.1

Specific surface was determined by retention of ethylene glycol monoethyl ether¹³.

Received 15 August; accepted 14 November 1977.

- Wada, K. & Harward, M. E. *Adv. Agron.* **26**, 211–260 (1974).
- Syers, J. K., Harris, R. F. & Armstrong, D. E. *J. environ. Qual.* **2**, 1–14 (1973).

3. Chen, B. H. H., King, P. H. & Randall, C. W. *Environ. Lett.* **6**, 129–138 (1974).
4. Eckenfelder, W. W. in *Water Quality Engineering for Practicing Engineers* (ed. Eckenfelder, W. W.) 220–224 (Barnes and Noble, New York, 1970).
5. Swenson, R. M., Cole, C. V. & Sieling, D. H. *Soil Sci.* **67**, 3–22 (1949).
6. Struthers, P. H. & Sieling, D. H. *Soil Sci.* **69**, 205–213 (1950).
7. Nagarajah, S., Posner, A. M. & Quirk, J. P. *Nature* **228**, 83–84 (1970).
8. Chen, Y. S., Butler, J. N. & Stumm, W. J. *Colloid Interface Sci.* **43**, 421–436 (1973).
9. Ng Kee Kwong, K. F. & Huang, P. M. *Clays Clay Miner.* **23**, 164–165 (1975); *Soil Sci. Soc. Am. J.* **41**, 692–697 (1977).
10. Gallet, J. P. & Paris, R. A. *Anal. chim. Acta* **39**, 341–348 (1967).
11. Toy, A. D., Smith, T. D. & Pilbrow, J. R. *Aust. J. Chem.* **26**, 1889–1892 (1973).
12. Hsu, P. H. in *Minerals in Soil Environments* (eds Dixon, J. B. & Weed, S. B.) Ch. 4 (Soil Sci. Soc. Am. Inc., Madison, Wisconsin, 1977).
13. Eltantawy, I. M. & Arnold, P. W. J. *Soil Sci.* **24**, 232–238 (1973).
14. Greenland, D. J. *Soil Sci.* **111**, 34–41 (1971).

Limited response of the K–Ar system to the Nordlinger Ries giant meteorite impact

For the proper interpretation of radiometric rock ages it is important to know which processes are able to reset radioactive clocks. In particular, the resetting by intense shock effects is significant in lunar chronology because most lunar highland rocks are impact breccias produced in the course of multiple large meteorite impacts. For the important K–Ar scheme, we have dated variously shocked samples from ejecta and from a drill core through one of the best documented terrestrial meteorite craters the Nordlinger Ries (Southern Germany)¹. From K–Ar and fission track dating of thoroughly melted impact glasses the 24 km diameter Ries crater was found to have formed 14.7 ± 0.7 Myr ago by meteorite impact into the variscian bedrock (for summary of ages see ref. 2). The modern version of K–Ar dating (^{39}Ar – ^{40}Ar technique) can reveal quantitative information about partial losses of radiogenic argon (fractional resetting) in addition to yielding ages eventually not affected by these gas losses^{3,4}. We have applied this dating technique to mineral separates (hornblende, biotite, chlorite) from ejecta, fall-back breccias and underlying bedrock of the Ries Crater, including drill core samples taken at depths up to 1,201 m (ref. 1). The samples exhibit various degrees of mechanical strain (from ~ 10 kbar up to > 400 kbar shock pressure). Apart from the thoroughly molten glasses which reproduced the impact age of ~ 14.7 Myr, we found that all other rocks involved in the impact irrespective of the degree of shock metamorphism and within the limits of error, yield the same ^{39}Ar – ^{40}Ar age of 313 ± 3 Myr, probably the age of the bedrock. Apparently, rocks exposed to the very intense shock effects frequent on the Moon give reliable K–Ar ages when measured as K–Ar plateau ages. Age resetting, if present, probably results from heating associated with such impacts. In the case of the Ries, we found from the diffusion characteristics of argon that the post-shock temperature of the crystalline fragments within the suevite layer has not exceeded 600°C .

Mineral separates were used to study the effect of shock on different minerals. We separated hornblende, biotite, and chlorite from the low grade metamorphic amphibolites and gneisses which occur as crystalline fragments imbedded in the suevite fall-back breccia above the crater bottom at 602 m depth, and as massive crystalline rocks below the crater bottom as well as in the crystalline wall of the crater (sample Meyer's Keller)¹.

The crystalline rocks from the drill core are weakly shocked and show impact features (like, for example, planar elements) typical of shock pressures between 10 and 200 kbar (stage '0' and '1' of shock metamorphism)^{5–7}. The crater wall amphibolite (Meyer's Keller) and the gneiss from ejected suevite sampled at Otting display shock effects of stage 2 (~ 300 – 450 kbar). All samples dated are listed in Table 1, including a melted impact glass and a moldavite that is, a tektite produced in the Ries impact event.

The experimental procedures applied in the ^{39}Ar – ^{40}Ar analysis are basically the same as those applied by us in dating lunar and meteoritic samples⁸. The samples obtained a fast neutron irradiation (FR2-reactor, Karlsruhe) to convert a part of the ^{39}K to the ^{269}yr – ^{39}Ar as a measure of the K in the sample.

Afterwards, the samples were degassed by stepwise heating from 400 to $1,500^\circ\text{C}$ (6–12 steps) and the Ar isotopes (^{36}Ar – ^{40}Ar) were measured by mass spectrometry. The ^{40}Ar to ^{39}Ar ratios obtained for each temperature fraction were converted into K–Ar ages by means of a muscovite monitor (Bern M4, $t = 18.0 \pm 0.2$ Myr) (ref. 9) irradiated together with the sample.

Two typical age spectra are shown in Fig. 1. In the absence of Ar-diffusion losses, an age plateau is obtained throughout essentially the whole range of degassing temperatures (Fig. 1a, hornblende from the crater bottom), whereas with the Ar-losses the geologically significant age plateau starts at intermediate temperatures and the low temperature fractions have reduced apparent K–Ar ages (Fig. 1b) ejecta from Otting.

A summary of results is given in Table 1. The character of the age spectra can be visualised from the error of the plateau age, the plateau range, and the percentage of gas loss. With the exception of the impact glass and the moldavite, both of which gave the impact age, 14.7 Myr, all samples yield plateau ages of approximately 313 Myr. The samples included depths > 700 m where the effect of the impact is minimal¹. The average age 313 ± 3 Myr is probably the age of the basement rock. The age obtained represents the first direct dating of the pre-Ries basement rock.

The uniformity of plateau ages from all depths (above and below the crater bottom) suggests that the impact process involving enormous amounts of energy¹⁰ did not effectively disturb the K–Ar system of hornblende or biotite. This is true for ejected material as well as for crystalline inclusions in both high and low temperature suevite, and agrees with the experimental results of Davis¹¹ who demonstrated in laboratory experiments (65–270 kbar) that shock pressure alone has little influence on the K–Ar system.

For natural impacts, fractional age reductions have been observed at shock pressures above 200 kbar for anorthositic

Fig. 1 ^{40}Ar to ^{39}Ar age spectra for hornblende from the crater bottom (a, H 603) and for suevite ejecta sampled at Otting (b, HO). Each data bar represents one release fraction in the incremental heating procedure, beginning at 450°C at the left and ending with 1500°C at the right end. The amount of gas released in the various fractions is reflected in the widths of the bars, expressed as proportion of the total ^{39}Ar contained in the sample. The measured ^{40}Ar to ^{39}Ar ratios are transformed into apparent K–Ar ages using a monitor irradiated together with the samples. H 603 exhibits a nearly perfect age plateau comprising 94% of the total argon release. Sample HO has lost 12.9% of its radiogenic argon from the less retentive sites (lower degassing temperatures), but the plateau age remains undisturbed and is properly defined by the higher temperature fractions, in spite of a shock pressure exceeding 400 kbar.

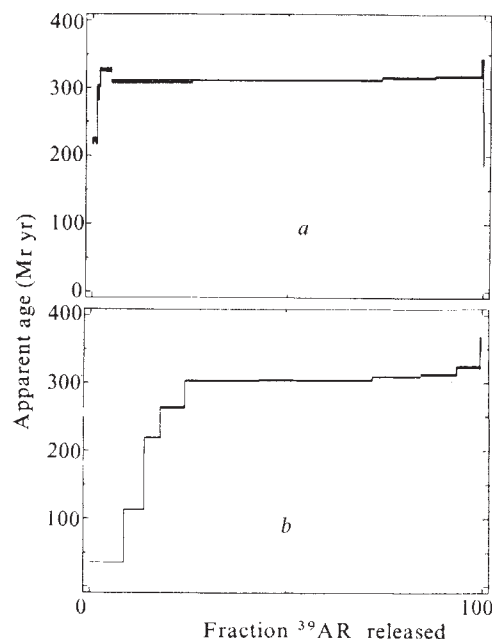


Table 1 Summary of ^{40}Ar – ^{39}Ar results for Ries crater samples

SAMPLE		K	Total $^{40}\text{Ar}_{\text{rad}}$	Total $^{40}\text{Ar}_{\text{atm}}$	Total K–Ar–age	Plateau age	Plateau range	^{40}Ar loss	Shock pressure
Position*	Parent rock	[%]	[$10^{-8}\text{cm}^3\text{g}^{-1}$]	[%]	[Myr]	[Myr]	[%–%]	[%]	[kbar]
Near surface samples									
HMK	Amphibolite	0.50	692	3.3	311 ± 11	315.0 ± 10.6	3–98	0.5	300–400
HO	Gneiss, suevite ejecta	0.51	675	11.0	268 ± 5	304.4 ± 4.8	24–94	12.9	400–450
BO	Gneiss, suevite ejecta	4.15	5,200	3.3	290 ± 6	316.6 ± 1.9	24–100	8.9	400–450
Drill core samples above the crater bottom									
B 377	Suevite breccia	1.89	2,421	4.2	296 ± 3	316.3 ± 2.4	22–100	7.0	n.d.
H 401		0.92	1,181	3.5	269 ± 3	302.4 ± 3.0	16–75	5.8	< 100
B 401	} Amphibolite, Suevite breccia	3.61	4,514	30.1	290 ± 5	not defined	—	—	< 100
C 401		0.50	625	28.9	287 ± 7	319.3 ± 3.4	28–100	10.8	< 100
B 503	Gneiss, suevite breccia	3.95	4,322	2.2	256 ± 5	315.4 ± 2.0	59–100	20.1	200 ± 50
H 570	Gneiss, suevite breccia	0.70	1,059	3.6	316 ± 10	316.6 ± 10.5	3–100	0.2	100–150
H 585	Gneiss, suevite breccia	0.71	881	5.1	263 ± 6	313.1 ± 5.4	18–98	16.2	$\lesssim 100$
Drill core samples below the crater bottom									
H 603	Amphibolite, crystalline	0.32	462	3.7	310 ± 5	310.9 ± 4.9	5–99	0.3	$\lesssim 100$
H 637	Amphibolite, crystalline	0.81	1,134	4.2	309 ± 3	310.5 ± 2.8	2–92	1.1	$\lesssim 100$
H 1201	} Gneiss Crystalline rock	0.67	949	3.5	312 ± 4	314.9 ± 3.7	9–99	1.2	< 10
C 1201		0.77	1,106	7.7	313 ± 15	316.9 ± 6.2	3–100	—	< 10
Glasses									
GO	Suevite breccia	2.68	166	30.0	14.6 ± 0.3	14.59 ± 0.49	3–92	—	melted
GM	Tektite (moldavite)	2.76	160	22.8	14.8 ± 0.14	14.82 ± 0.15	12–100	1.2	melted

*H = hornblende; B = biotite; C = chlorite; G = glass; MK = Meyer's Keller, crater wall; O = Otting; numbers indicate the depth in metres within the drill core.

Typical sample weights are 200–500 mg, except B 401 (5 mg). The K-content (error $\pm 2.5\%$) follows from the total reactor produced ^{39}Ar and the monitor derived conversion factor. Atmospheric ^{40}Ar is subtracted on the basis of ^{36}Ar (its relative magnitude is given in % of total ^{40}Ar). The total ('conventional') K–Ar age is derived from the integral $^{40}\text{Ar}_{\text{rad}}$ and ^{39}Ar quantities for all temperature fractions. In case of gas loss, it is lower than the plateau age. The error of the plateau age is the standard mean deviation of the ages of the plateau fractions. 'Plateau range' indicates the extension of those fractions which define the plateau age. These entries allow one to visualise the age spectra without explicit figure, examples for HO and H 603 are shown in Fig. 1. The shock pressure is estimated from microscopic inspection following the criteria of Engelhardt *et al.*⁶. Note the absence of a correlation between gas loss and shock pressure. Within the limits of error, essentially all samples yield the same age of $\approx 313 \pm 3$ Myr except the thoroughly molten glasses GO and GM which reproduced the impact age of 14.7 Myr.

from the 30 km Manicouagan structure (conventional K–Ar dating) (ref. 12) and for gneiss and feldspar from the 4 km Brent Crater (conventional K–Ar dating¹³, ^{39}Ar – ^{40}Ar dating¹⁴). This contrasts with the Ries observation of unreduced bedrock ages up to an above 400 kbar (Otting) not only for resistant hornblende but also for biotite. At both the Manicouagan and Brent Craters, samples from a recrystallised melt layer were studied, and probably means that age reductions observed for Manicouagan and Brent Craters are related to a thermal event. From the Ries K–Ar ages we conclude that shock pressure *per se* even as high as 400 kbar does not cause any significant argon loss. Consequently, any eventual age changes caused by meteorite impacts must be related to the temperature elevations associated with the impact and the rate of heat dissipation afterwards.

The peak temperature for the Ries fall-back breccia may be estimated. The common onset temperature, 800 °C, of the age plateaux of crystalline inclusions indicates that the inclusions were never heated above that temperature for any appreciable lengths of time ($\sim$ days). This upper limit temperature can be reduced to 600 °C by a model calculation for the cooling history of the suevite layer and an estimate of the hornblende activation energies from the stepwise heating argon release pattern. A 500 °C lower limit has been inferred by Wagner¹⁵ from the thermal stability and the degree of annealing of U-fission tracks in titanite and apatite from the ejected suevite from Otting. Apparently, the Ries ejecta was heated to 550 ± 50 °C at the time of impact.

This work was supported by Gesellschaft für Kernforschung mbH, Karlsruhe.

E. K. JESSBERGER
TH. STAUDACHER
B. DOMINIK
T. KIRSTEN
O. A. SCHAEFFER

Max-Planck-Institut für Kernphysik,
6900 Heidelberg,
FRG

Received 7 November; accepted 23 November 1977.

1. *Geologica bav.* 75, 1–470 (1977).
2. Gentner, W. & Wagner, G. A. *Geologica bav.* 61, 296–303 (1969).

3. Merrihue, C. M. & Turner, G. J. *geophys. Res.* 71, 2852–2857 (1966).
4. Kirsten, T. in *The Origin of the Solar System* (ed. Dermott, S. F.) (in the press).
5. Engelhardt, W. von & Graup, G. *Geologica bav.* 75, 255–271 (1977).
6. Engelhardt, W. von, Stöffler, D. & Schneider, W. *Geologica bav.* 61, 229–296 (1969).
7. Chao, E. C. T. *Neues Jb. Miner. Abh.* 108, 209–246 (1968).
8. Jessberger, E. K., Kirsten, T. & Staudacher, Th. *Proc. Lunar Sci. Conf.* 7, 2201–2215 (1976).
9. Jaeger, E. *Colloquium on the Geochronology of Phanerozoic Orogenic Belts*, Zurich and Bern (1969).
10. Chao, E. C. T. *Geologica bav.* 75, 421–441 (1977).
11. Davis, P. K. *Geochim. cosmochim. Acta* 41, 199–206 (1977).
12. Wolfe, S. H. J. *geophys. Res.* 76, 5424–5436 (1971).
13. Hartung, J. B., Dence, M. R. & Adams, J. A. S. J. *geophys. Res.* 76, 5437–5448 (1971).
14. Jessberger, E. K. *et al. Conf. Comp. Mercury, Moon, Lunar Sci. Inst.*, Contr. 262, Houston, Texas (1976).
15. Wagner, G. A. *Geologica bav.* 75, 349–354 (1977).

Alternative ^{210}Pb dating: results from the New Guinea Highlands and Lough Erne

RECENT lake sediments can be dated using ^{210}Pb and fall-out ^{137}Cs . Pennington *et al.*¹, and Robbins and Edgington² have set out the assumptions used in calculating dates and estimating accumulation rates from the declining concentration of unsupported ^{210}Pb in the near-surface sediments of Blelham Tarn and Lake Michigan respectively. In both cases, as in other papers^{3,4}, an essential assumption is a constant initial concentration (c.i.c.) of unsupported ^{210}Pb per unit dry weight in the sediment at each depth, whether or not any variations may have occurred in the rate of accumulation. This assumption requires that in undisturbed cores, unsupported ^{210}Pb concentrations should always decline monotonically with depth. Figure 1 shows unsupported ^{210}Pb concentrations in cores from Lough Erne, Northern Ireland and Lake Ipea, Papua New Guinea. The profiles are 'kinked' and show at one or more points, a marked increase in unsupported ^{210}Pb concentration with depth. The levels at which this occurs in the cores range from 6 to 30 cm. The increases cannot, therefore, be the result of the anomalously low surface concentrations noted elsewhere⁴. Associated biological, chemical and geophysical studies show that the profiles have not been significantly disturbed by physical or biological mixing. These profiles are not consistent with the c.i.c. deposition model and are

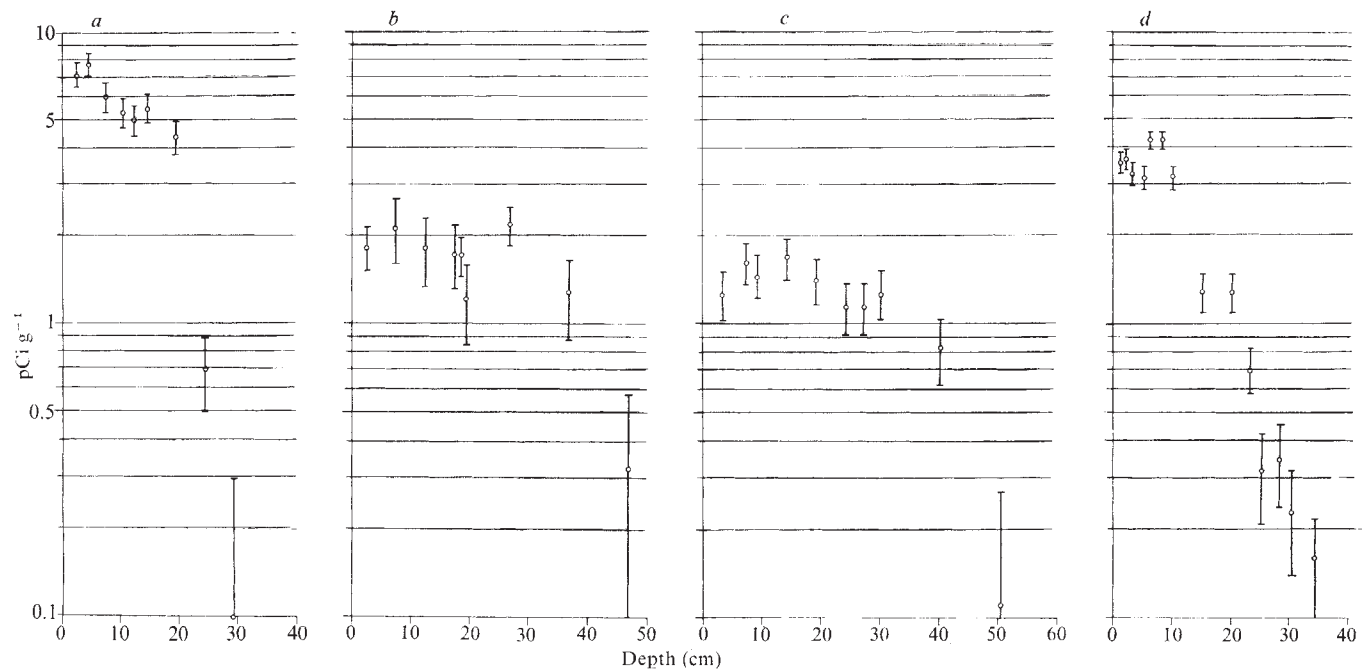
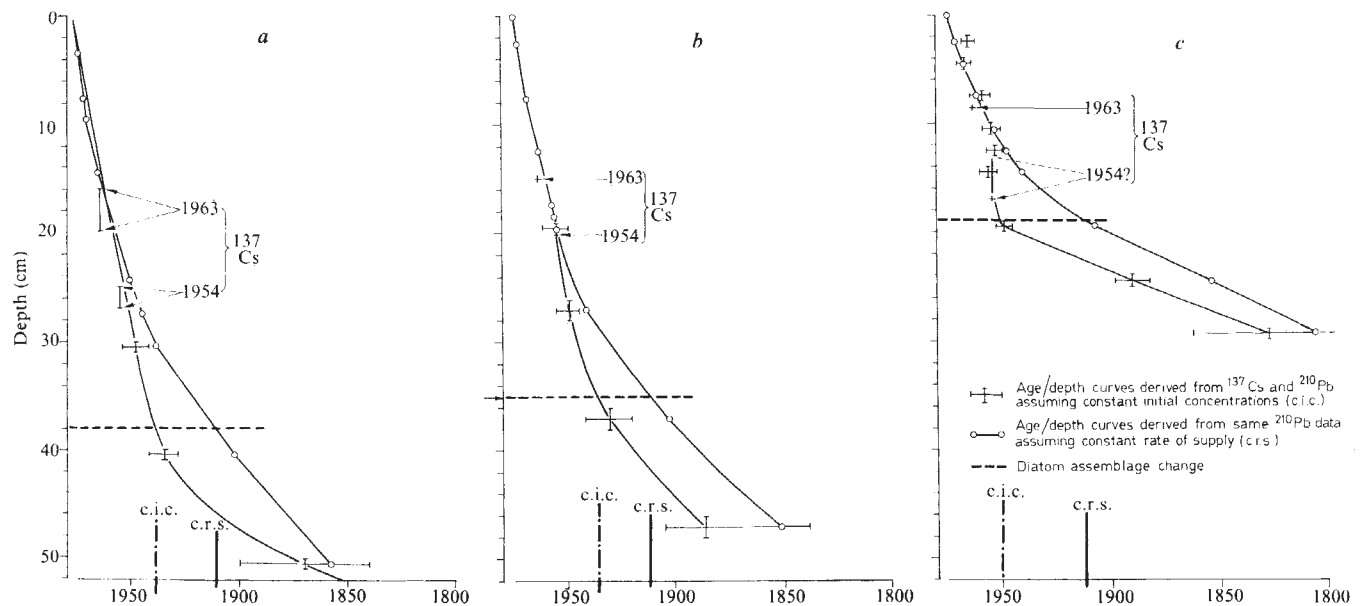


Fig. 1 Unsupported ^{210}Pb concentration against depth in cores from Lough Erne, Northern Ireland, and Lake Ipea, Papua New Guinea. *a*, Lower Lough Erne, profile SM1; *b*, Lower Lough Erne profile FM1; *c*, Upper Lough Erne profile FM2; *d*, Lake Ipea, Core 1.

regarded as evidence for the dilution of unsupported ^{210}Pb by accelerated sediment accumulation. If such dilution has taken place without leading to a kink in ^{210}Pb concentrations, the assumption of c.i.c. will lead to underestimation of the true age of the sediment below the onset of acceleration. In the case of the 'kinked' profiles, dates are not calculable using the c.i.c. deposition model alone. We have adopted an alternative approach to calculating ^{210}Pb dates using as our main assumption a constant rate of supply (c.r.s.) of unsupported ^{210}Pb to the sediment per unit time and deriving dates from the integrated activity of the radionuclide. Previous authors have referred to the possibility

of calculating dates using this assumption^{5,6} but have not given a full account of the method or evaluated the results of its application. Details of the methods used here are set out elsewhere⁷. This paper compares and briefly evaluates the two alternative models as applied to the sediments of Lough Erne and Lakes Ipea and Egari. The c.r.s. based dates obtained have been compared with those derived either from c.i.c. based ^{210}Pb dates or, in the case of the 'kinked' profiles, from a combination of ^{137}Cs dating⁸ and c.i.c. based calculations. Figure 2 plots the resulting age against depth curves from Lough Erne, Fig. 3, some of the results from Lakes Ipea and Egari.

Fig. 2 Alternative ^{210}Pb based age against depth curves from three cores from Lough Erne, Northern Ireland. The standard errors are shown for the c.i.c. based profiles only. For cores *b*, FM1 and *a*, FM2 the strongly 'kinked' unsupported ^{210}Pb concentration profiles above 30 cm (see Fig. 1) make it impossible to derive c.i.c. based age against depth curves above this level. *c*, SM1. The alternative dates for the diatom assemblage change are marked on the timescale at the base of each plot.



Diagrams of diatom frequency have been completed for all three Lough Erne cores. The main change in diatom assemblages in each core is marked on Fig. 2. In each case, at this stage in the core, the proportion of planktonic diatoms expands and the concentration of frustules (cm^{-3} fresh weight) begins to increase. Using the ^{137}Cs and c.i.c. based age/depth curves, the age of this change ranges from ~ 1935 in core FMI to ~ 1950 in core SM1, both from the Lower Lough. Using the c.r.s. model, dates in all three cores lie between 1907 and 1914.

On the basis of both calculations, sediments at 24.5 cm in SM1 and at 47 cm in FMI seem to be contemporary. It is therefore possible to normalise these two profiles approximately, with regard to age. The c.i.c. model would require either that the ratio of concentrations at contemporaneous depths in the two sites to be constant or, because the dry weight accumulation rate above this level in FMI is over three times as fast as in SM1, that the cumulative residual unsupported ^{210}Pb values in FMI be greatly in excess of those in SM1. The ratio of concentrations between FMI and SM1 varies from 2.3 : 1 to 4 : 1. Moreover, Fig. 4 shows that

each lake and, in both cases, the first schemes were installed around the turn of the century.

The presence of a synchronous volcanic ash layer in the recent sediments of Lakes Ipea and Egari makes it possible to identify two points at which the sediments for the two lakes are exactly contemporaneous—the surface mud and the mud in contact with the volcanic ash. Assuming a c.i.c. of unsupported ^{210}Pb implies that for all stages of deposition known to be synchronous between the two sites, the ratio between the concentrations of unsupported ^{210}Pb at the two lakes must be the same. Figure 5 shows that this is not the case; for example, in recent times the concentration at Egari has been around three times that at Ipea whereas the ratio is close to one during the period just after the ash fall. Though not compatible with the c.i.c. model this evidence is consistent with a c.r.s. model in which unsupported ^{210}Pb is diluted by more rapidly accelerating accumulation at Ipea than at Egari.

Geochemical studies (Blong, personal communication) have confirmed that the source of the most recent ash fall recorded in the Ipea and Egari sediments, as well as exten-

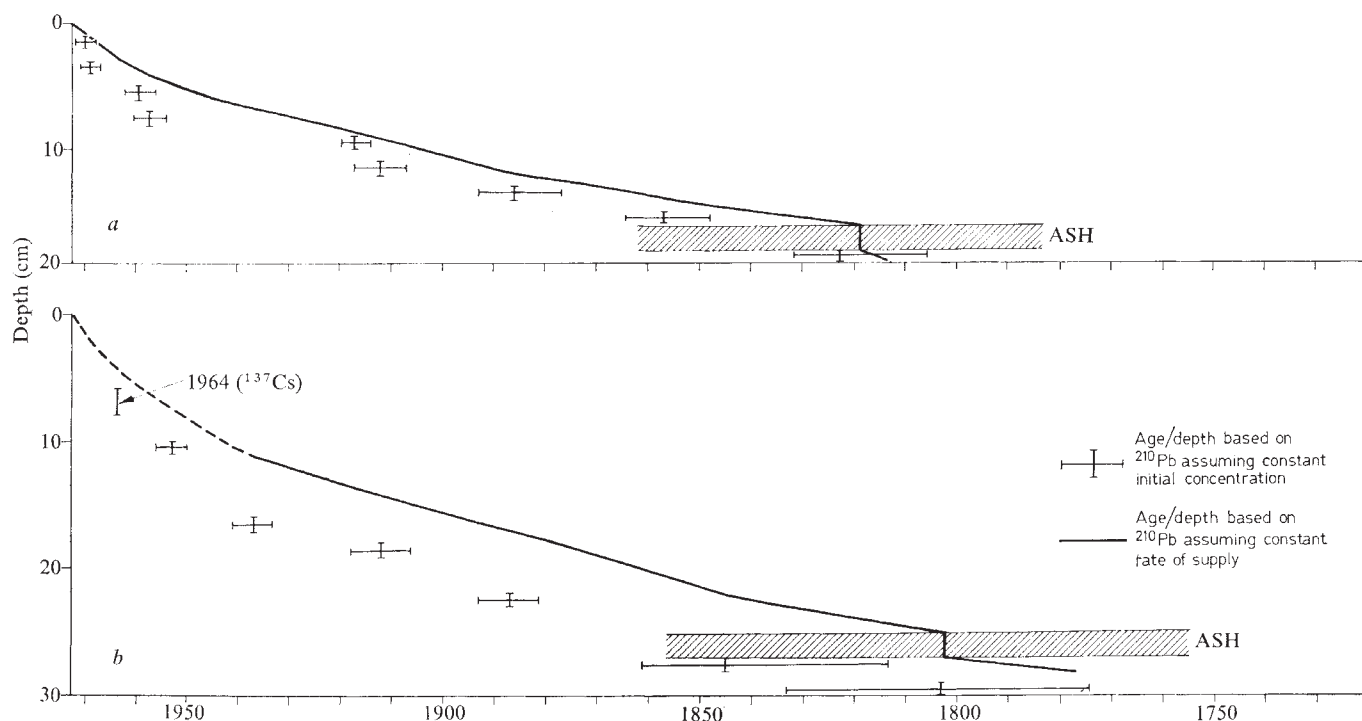


Fig. 3 Alternative ^{210}Pb based age against depth curves for *a*, core 5 from lakes Egari; and *b*, core 4 Ipea in the Highlands of Papua New Guinea. The ^{137}Cs age estimate takes into account the lag in peak fall-out between hemispheres. The upper peaked section of the c.r.s. based age against depth curve from Ipea has been obtained from measurements on core I for which detailed recent data are available; below this level, the data from the parallel adjacent core 4 have been used since at the lowest levels the ^{210}Pb measurements for core I are somewhat high when compared with mutually consistent profiles from the two sites.

the curves and cumulative residual unsupported ^{210}Pb values are almost identical and hence consistent with the c.r.s. model in which the rate of supply of unsupported ^{210}Pb to each site has been comparable, despite the difference in the rate of sediment accumulation.

No firm independent evidence of age is available for any of the recent Lough Erne sediments. The change in the diatom record noted above and redated to around 1910 indicates increasingly eutrophic conditions. It parallels that also dated to the early part of the twentieth century in Lough Neagh⁹. The most likely cause of this is the installation of sewerage schemes within the drainage basins of

sively elsewhere in the New Guinea Highlands, was Long Island some 400 km to the east. Analysis of historical records has shown that the period during which a major eruption in Long Island could have taken place ended long enough before 1827 to have left no obvious major traces on the landscape of the island¹⁰. Since then, the island has been frequently visited and its state well documented by Western observers. Several radiocarbon dates on terrestrial material associated with the ash placed its age at $\sim 250 \text{ BP}^{11}$. Both lines of evidence point to an eighteenth century date and they are quite incompatible with the original c.i.c. based estimates which averaged $\sim \text{AD}1860$. The four recalculated

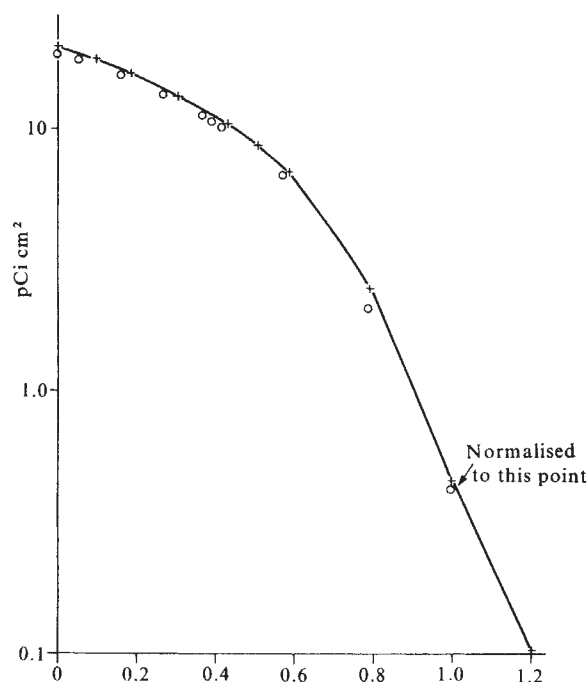
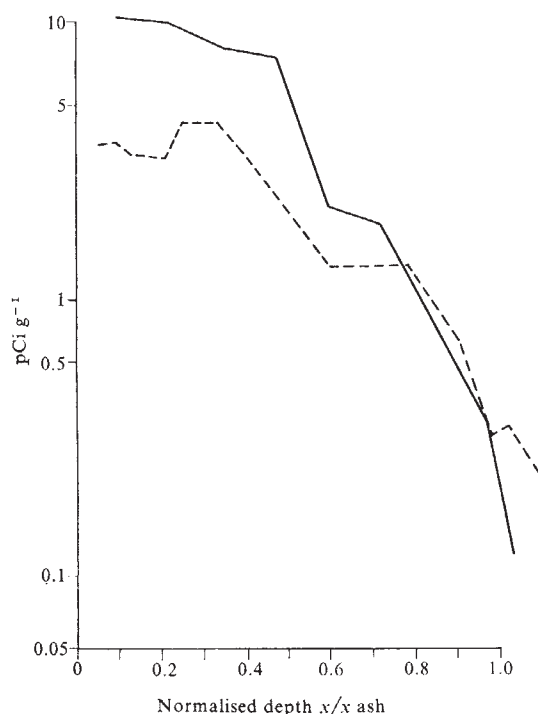


Fig. 4 Cumulative total residual unsupported ^{210}Pb content for Lower Lough Erne cores FM1 (○) and SM1 (+) normalised to a correlated depth.

^{210}Pb profiles from Lakes Ipea and Egari regarded as mutually consistent on internal evidence give an age range between 1800 and 1840 and an average age of 1814. This date is within the time range possible on documentary grounds and much closer to the ^{14}C dates than were the original ^{210}Pb estimates.

Where accumulated rates have not changed over the last 150 years both the c.i.c. and the c.r.s. models will give identical and possibly reliable dates. The data presented

Fig. 5 Unsupported ^{210}Pb concentrations from Egari core 5 (solid line) and Ipea core 1 (dashed line) normalised for depth (×) by the synchronous volcanic ash layer common to both sediment profiles.



here suggest that in a lake where accumulation rates have accelerated, the c.r.s. model may provide dates which are more consistent on internal grounds and more compatible with existing external evidence.

We thank NERC, the Nuffield Foundation and the Department of the Environment, N. Ireland for financial assistance. We also thank J. D. Eakins and R. S. Cambray for the ^{210}Pb and ^{137}Cs profiles and R. Blong for encouragement.

F. OLDFIELD

Department of Geography,

P. G. APPLEBY

Department of Applied Mathematics
and Theoretical Physics,
University of Liverpool, UK

R. W. BATTARBEE

Department of Geography,
University College,
London, UK

Received 8 June; accepted 8 November 1977.

1. Pennington, W., Cambray, R. S., Eakins, J. D. & Harkness, D. D. *Freshw. Biol.* 6, 317 (1976).
2. Robbins, J. A. & Edgington, D. N. *Geochim. cosmochim. Acta* 39, 285 (1975).
3. Krishnaswamy, S., Lal, D., Martin, J. M. & Maybeck, M. *Earth planet. Sci. Lett.* 11, 407 (1971).
4. Koide, M., Bruland, K. W. & Goldberg, E. D. *Geochim. cosmochim. Acta* 37, 1171 (1973).
5. Goldberg, E. D. *Symp. Radioactive Dating* 121, (IAEA, Vienna, 1963).
6. Crozaz, G., Picciotto, E. & De Breuck, W. J. *Geophys. Res.* 69, 2597 (1964).
7. Appleby, P. G. & Oldfield, F. *Catena* (in the press).
8. Pennington, W., Cambray, R. S. & Fisher, E. M. *Nature* 242, 324 (1973).
9. Battarbee, R. W. *Phil. Trans. Roy. Soc., Lond. B.* (in the press).
10. Ball, E. & Johnson, R. W. in *Volcanism in Australasia* 133 (Elsevier, Amsterdam 1976).
11. Blong, R. J. J. *Polynes. Soc.* 84 (1975).

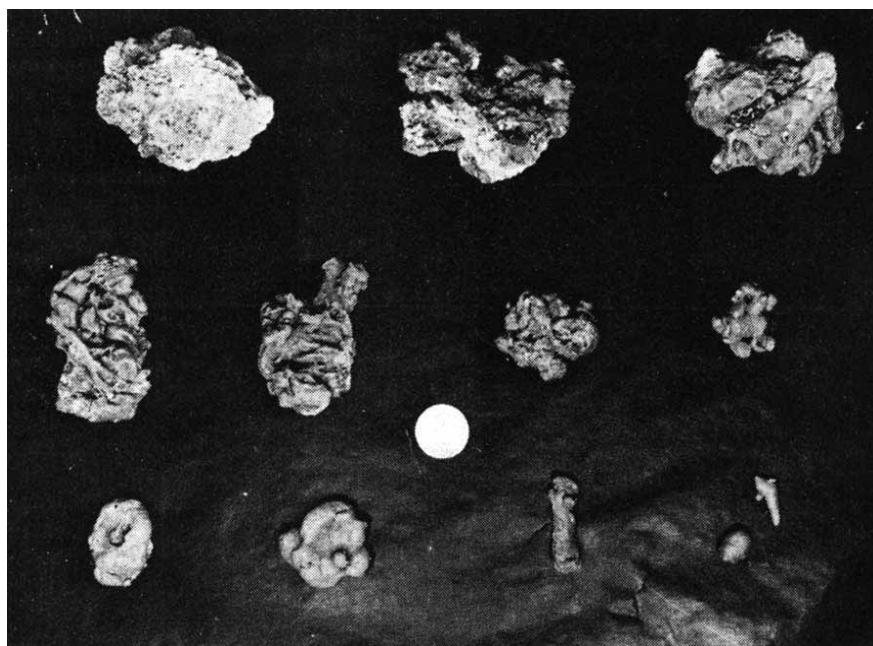
Subaqueous sulphur lake in Volcan Poas

GREENISH amorphous sulphur particles are present at all levels immediately around the crater of Volcan Poas. The eruptive plumes of the mild submarine volcanism which characterises this volcano, are charged with these particles. We suggest here that they are strong evidence that a lake of molten sulphur exists, or is formed temporarily before an eruption, below the existing lake of water in the crater. We have collected many samples of greenish amorphous sulphur, with various quantities of admixed ash from regions all around the crater, noting some concentrations in the east-west prevailing wind directions. Figure 1 shows sulphur specimens and their morphology; Fig. 2 gives a profile of the crater; Fig. 3 shows approximate outlines of areas in which sulphur samples are found. Table 1 gives approximate analyses of representative samples.

Volcan Poas has been in a state of mild eruptive activity for about two centuries. Its usual activity is similar to shallow submarine volcanism. At intervals ranging from several times a month to several times an hour a plume arises from the approximate centre of the crater lake and ascends to heights varying from a few metres to 500 m or more. The plume is laden with particulate matter and may appear black or greenish. At much longer intervals, more violent eruptions producing grey ash, scoria, plastic lavas and bombs will occur. The most recent of these occurred in November 1976, and spread a mantle of grey ash for some kilometres to the west of the crater.

We shall look at the plume eruptions and their similarity to submarine volcanism, and we shall explain them using an approach¹ which may be generally applicable. All active volcanic structures contain a stack, or pile, of immiscible liquids separated by more or less definite boundaries; within each level processes of heat and mass transfer may occur by conduction and convection; bubbles may form, grow,

Fig. 1 Samples of amorphous sulphur from Volcan Poas. Ash content high in sample at upper left. Coin diameter = 1.5 cm.



rise and attach to upper boundaries, or disappear; processes of dispersion, sedimentation and crystallisation develop within layers. Across boundaries, or in boundary layers, fluxes of mass, heat, bubbles, crystals, and so on may occur. When the local vapour pressure within a particular level exceeds the hydrostatic pressure at that level, and the temperature distribution above will permit the growth of a vapour-filled cavity, then a suitable disturbance may initiate the vaporisation expansion wave which supports the growth of a plume, and an explosive eruption will take place.

In the present example, we consider the volcanic pile to be similar to that for shallow submarine volcanism. This assumption implies a volcanic pile with dry magmas below, more volatile magmas above these, then a crust or plastic layer, water-rich muds and clays, and water suspensions grading from heavily laden to lightly laden in the upward direction.

In Fig. 1 the sulphur samples all show evidence of a fluid dynamic eruptive process. The plume gases entrain and interact with molten or plastic, liquid sulphur by means of the reverse flows around the deformable fluid particles being accelerated by the vertical jet in the plume. Representative particles of the collection show elongations, separations into branches with ball-like terminations, twisted streamline configurations, deformed spheres, tiny spherical droplets, plastic masses with considerable ash content; and there exist some of all these types with vesiculated surfaces and interiors. These latter were probably burning during the eruption, over at least part of their trajectories.

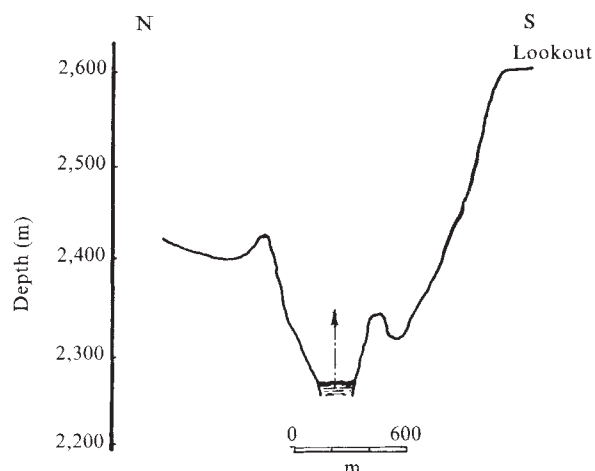
Experiments in which existing pieces of our collection are set on fire show a clear blue flame characteristic of sulphur, produce SO_2 fumes, and leave very little residual ash. We observe vesiculation both of the surface and of the interior under the flame.

According to accounts of the rangers at Volcan Poas National Park, bushes at the high plateau level, 300–500 m, above the lake, were set on fire. We take the region devoid of living vegetation as the area covered, during various eruptions, by sulphur particles. In a circle of radius 1/2 mile around the lake centre, the density of sulphur particles suggests that several hundred tons of sulphur are lying on the surface or have been washed into the river courses by the heavy rains of the wet season.

The distribution with height, and the change from high sulphur content at lower altitudes, to less sulphur and more ash at the upper altitudes above the lake suggests that the distribution in the columnar plume reflects a similar distribution in sulphur content of the portion of the volcanic pile where sulphur occurs.

We believe that the evidence available justifies the assumption that there is a level, or lake, of molten sulphur below the existing lake of water. The specific gravity of sulphur, 1.9–2.0, is intermediate between that of andesitic magma, 2.5–2.7, and that of pure water. We have no evidence for the average specific gravity of a water-rich clay or mud of volcanic ash, but we assume that values intermediate in the range 1.0–2.0 can be encountered. Therefore, to our model of the volcanic pile we add, immediately above the crusted magma, a layer of molten sulphur on which floats a layer of water-rich volcanic ash and lapilli. Above this layer lies the crater lake which from its greenish appearance carries a suspended load of sulphur and ash varying with time. One of the precursor signs of an imminent eruption is the appearance of a grayish spot, 25–50 m in diameter, in the centre of the lake. We take this to be

Fig. 2 Profile of the crater. Arrow indicates the position of the plume.



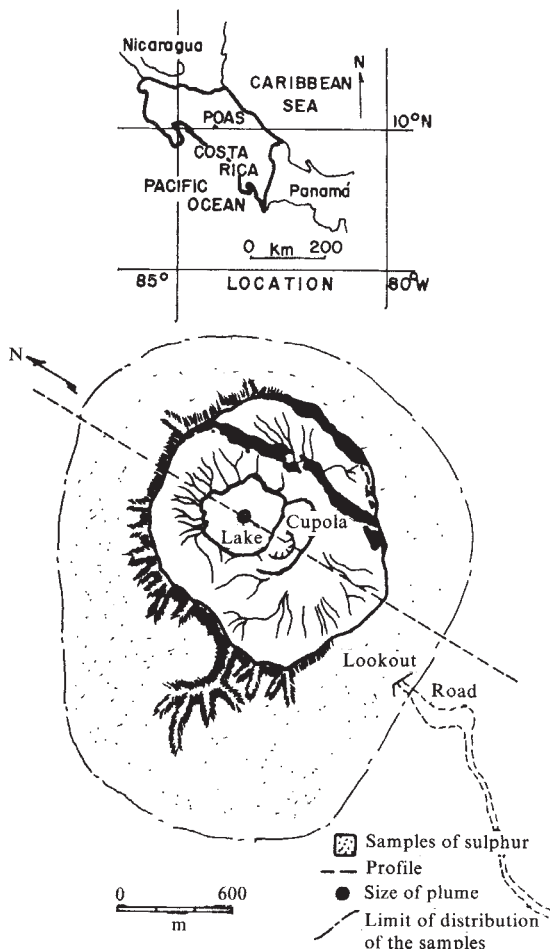


Fig. 3 Map of the crater.

an indication of convective processes that carry portions of the underlying sulphur-ash levels to the surface.

We can only speculate how the plume forms and the eruption occurs. Because the plume is laden with molten sulphur-ash particles some of which catch fire on exposure to air, we may presume high temperatures in the sulphur layer, probably higher than the normal boiling point, 444 °C, in the lowermost portions. If the propulsive fluid for the eruption is sulphur vapour, then temperatures much higher than the boiling point will be encountered at the level between solid or plastic magma and molten sulphur.

A possible mechanism for production of an eruption, other than the gradual increase in temperature of the column during the hydrostatic phase, would be an overturning instability brought about by an increase in density of the mud deposited by sedimentation from above. As the density of the overlying mud increases to near equality with that of the molten sulphur, an overturning instability would be possible that would rapidly expose high temperature sulphur to overlying water and bring about an explosion similar to those noticed in fuel-coolant interactions³.

Table 1 Analysis of sulphur content of representative samples

Sample	1	2	3	4	5
Weight % sulphur	16	85	98	98+	99

We think that other sulphur producing volcanoes, particularly those of andesitic type, should be examined for signs of the existence of a lake of molten sulphur in the volcanic pile.

S.M.R. is a visiting professor in the cooperative programme between the governments of Italy and Costa Rica.

F. D. BENNETT*

S. M. RACCICHINI

*Escuela de Ciencias Geográficas,
Universidad Nacional,
Heredia, Costa Rica*

Received 19 August; accepted 1 December 1977.

* Present address: Department of Mechanics and Materials Science, The Johns Hopkins University, Baltimore, Maryland 21218.

1. Bennett, F. D. *J. geophys. Res.* **77**, 5755–5759 (1971).

2. Buchanan, D. J. *J. Phys. D: appl. Phys.* **7**, 1441–1457 (1974).

The Soufrière Crater Lake as a calorimeter

IN 1971–72, $8 \times 10^7 \text{ m}^3$ of basaltic andesite lava were extruded into the summit crater of Soufrière volcano, St Vincent¹. The first half of the eruption took place entirely subaqueously and the second half almost entirely so and these facts have provided us with an unusual opportunity to carry out quantitative studies of several aspects of the eruption. Before the eruption the crater contained $7.5 \times 10^7 \text{ m}^3$ of water and the potential of this lake as a giant condenser of volcanic gases has already been demonstrated². We show here that by treating the crater lake as a giant calorimeter we can make reasonably accurate estimates of the rate at which heat was lost by the erupted lava and hence draw important conclusions about its mode of extrusion. The heat transfer data are consistent with the idea that the lava mass accumulated by the multiple extrusion of numerous small lava bodies¹ rather than by extrusion of a single domelike mass, as has been proposed by other workers^{3, 4}.

The eruption of 1971–72 began some time in early October 1971 and until November 20, 1971 it continued entirely subaqueously. Before the eruption the crater lake contained $7.5 \times 10^{10} \text{ kg}$ of water at a temperature close to the mean ambient temperature of 22 °C. Seepage into the crater walls and normal evaporation were balanced almost exactly by precipitation. By the time that lava began to emerge through the surface of the lake on 20 November 1971, the mass of water in the lake had decreased to $5.7 \times 10^{10} \text{ kg}$. The temperature at the edge of the lake was 82 °C on 2 November and probably was maintained at at least this value for at least 1 month before 20 November. Throughout this early part of the eruption the temperature of the streams and springs on the outer flanks of the volcano was monitored both by aerial infrared photography and by manual measurements and we found no evidence that seepage from the lake had increased. We therefore conclude that $1.8 \times 10^{10} \text{ kg}$ of water were lost by excess evaporation caused by heat input from the lava. Heat supplied to the crater lake up to 20 November therefore includes four main components. (1) The heat required to raise $7.5 \times 10^{10} \text{ kg}$ of water from 22 °C to 82 °C ($1.9 \times 10^{16} \text{ J}$). (2) The heat required to raise $1.8 \times 10^{10} \text{ kg}$ of water to the boiling point at the elevation of the lake surface (98 °C) ($0.1 \times 10^{16} \text{ J}$). (3) The heat required to convert $1.8 \times 10^{10} \text{ kg}$ of water to steam at 98 °C ($4.1 \times 10^{16} \text{ J}$); and (4), the heat lost from the surface of the lake to the atmosphere by radiation and convection. The radiative heat-loss can be estimated from the Stefan-Boltzman equation

$$\delta H / \delta t = \sigma \Sigma (T_s^4 - T_o^4) W \text{ m}^{-2}$$

where σ is Stefan's constant, Σ is the emissivity of the water surface and T_s and T_o are the surface and ambient absolute temperatures respectively. Inserting appropriate values for the variables and assuming that the lake reached 82 °C 1 month ($2.5 \times 10^6 \text{ s}$) before 20 November the radiative heat loss is $0.1 \times 10^{16} \text{ J}$. Similarly the heat lost by convection can be estimated from the approximate expression (ref. 4):

$$\delta H / \delta t = 1.52 (T_s - T_o)^{1/3} W \text{ m}^{-2} (\text{°C})^{-1}$$

which gives a convective heat loss of $0.1 \times 10^{16} \text{ J}$. The total

heat supplied to the lake water between the beginning of the eruption and 20 November was therefore at least 6.3×10^{16} J. This value is certainly an underestimate since it excludes heat lost by conduction into the crater walls and by convective transfer to pre-existing groundwater in the crater walls.

The mass of lava erupted up to 20 November was 10^{11} kg. The available thermal energy of the lava includes the enthalpy of melting (latent heat) and the heat lost by the lava in cooling from its eruptive to its final temperature. Since the lava pile was still growing on 20 November it is clear that it retained a fluid core which was probably still at the eruptive temperature of approximately $1,050^\circ\text{C}$ whereas the outer surface of the lava could not have been at a temperature greater than 100°C . There was therefore a carapace of solidified, though perhaps extensively fractured, lava separating the fluid core from the lake water with a temperature difference of 950°C across the carapace. Any physically reasonable heat-transfer mechanism, either by conduction or by circulation of fluids, requires that the temperature gradient across the carapace should be at least quasilinear which implies that the mean temperature of the carapace was the arithmetic mean of the water and lava temperatures. If the mass of lava forming the carapace was M kg, the total amount of heat lost by the lava in solidifying and cooling from $1,050^\circ\text{C}$ to 575°C was

$$H = M[L + 475C]$$

where L is the latent heat of the lava and C is its specific heat capacity. Taking $L = 3 \times 10^5 \text{ J kg}^{-1}$ and $C = 10^3 \text{ J kg}^{-1} (^\circ\text{C})^{-1}$, then

$$H = 7.75 \times 10^5 M \text{ J}$$

Equating the heat lost by the lava to the heat gained by the lake we have

$$M = 8.1 \times 10^{10} \text{ kg}$$

That is, of the lava erupted up to 20 November over 80% had solidified and cooled to a mean temperature 475°C below its eruptive temperature by that date. We emphasise that this is a lower limit, because the heat supplied to the lake water is known to be an underestimate. A previous assessment of the heat budget² did not take into account all the factors considered here and was an even greater underestimate of the volume of cooled lava.

The same pattern continued during the rest of the eruption. When extrusion of lava stopped on 20 March 1972 a total of 2×10^{11} kg of lava had been erupted, and 5×10^{10} kg of water had evaporated. By calculations similar to those above we estimate that about 90% of the lava had solidified when the eruption ceased. This again is a lower limit since we were unable to make accurate estimates of the heat lost by radiation directly from the lava to the atmosphere after it emerged from the lake. These observations therefore make it clear that during the 1971–72 eruption of the Soufriere of St Vincent the lava solidified at about the same rate that it was erupted and that the solidified lava cooled considerably after crystallisation.

During the eruption, large quantities of heat were transferred rapidly from the lava to the lake water. The characteristic dimension of a typical lava cooling unit can be estimated from the observation that the lava was able to lose a significant proportion of its heat content to its surroundings within a time period which was less than 1 month. The thermal time constant τ of a body with characteristic dimension R metres and thermal diffusivity $\kappa \text{ m}^2 \text{ s}^{-1}$ is of order R^2/κ . Putting $\tau = 1$ month ($2.5 \times 10^6 \text{ s}$), $\kappa = 10^{-6} \text{ m}^2 \text{ s}^{-1}$ we estimate that the maximum possible characteristic dimension of the individual cooling units was in the range 1–5 m. This result clearly excludes the possibility that the lava mass was extruded as a viscous dome^{3,4} with a characteristic dimension of order 500 m. There seem to be two possible ways in which the intimate mixing of lava and water necessary to cause the observed rapid heat-flow could have been achieved. The first involves breaking the whole

lava pile into individual blocks with diameters less than 5 m. This possibility can be excluded simply on the grounds of visual observations at the time of eruption¹. There were indeed rock avalanches during the later stages of the eruption but these involve only a very small proportion of the erupted mass. The second possibility, which is supported by the visual observations, is that the lava pile grew by the successive extrusion of overlapping flow-units, each with a characteristic dimension (which in the case of flow is its thickness) of order 1–5 m.

The combined field observations¹ and calorimetric evidence show that the 1971–72 eruption was a quiet extrusion of a lava flow. The gross morphologic form of the lava mass was controlled by the shape of the container (crater) and the cooling effect of the surrounding lake water, and not by the viscosity of the lava. Other features of the eruption, particularly the lack of associated seismicity, the absence of explosions and the composition of the lava suggest that the viscosity of the magma was low and the results presented here show that the gross morphology of the lava pile is consistent with this conclusion.

This eruption marks the first documented extrusion of a lava flow in the Lesser Antilles island arc. Previous events of extrusion of minor lava masses within the Soufriere crater lake are, however, indicated by very sketchy historic accounts¹, suggesting episodes of minor lava extrusion at intervals between the major explosive eruptions of this volcano. The fact that the Soufriere is capable of quiet lava flow eruptions as well as violent explosive ones must be considered in any theory of its eruptive mechanism.

J. B. SHEPHERD

Seismic Research Unit, University of the West Indies,
St Augustine, Trinidad

HARALDUR SIGURDSSON

Graduate School of Oceanography,
University of Rhode Island, Kingston,
Rhode Island 02881

Received 22 August; accepted 30 November 1977.

1. Aspinall, W. P., Sigurdsson, H. & Shepherd, J. B. *Science* **181**, 117 (1973).
2. Sigurdsson, H. *J. Volcanol. Geotherm. Res.* **2**, 165 (1977).
3. Roobol, M. J. & Smith, A. L. *Bull. Volcanol.* **39**, 214 (1975).
4. Tomblin, J. F. *Smithsonian Institution Center for Short-Lived Phenomena, A. Rep.* 1971, 217 (1972).
5. Roger, G. F. C. & Mayhew, T. R. in *Engineering Thermodynamics* (Longman, London, 1972).

Circular structures of large scale and great age on the Earth's surface

THE Earth's surface exhibits faint circular patterns which have not been described before. These circles are characterised by near perfection of outline, by the presence of topographic highs (rims) along parts of their circumferences, by their generally large scale (diameters of from under 7 km up to approximately 700 km in the areas examined), and by their definition in various geological environments, in many rock types, and in rocks of all ages. Many of the circles are intermittent in places along their rims but about 55% of the approximately 1,170 definite circles observed to date can be visually traced around an entire 360° of arc. The circles are further characterised by the presence of fracturing and brecciation along parts of their rims and by the extraordinary control they place on regional geology in general and on ore mineralisation in particular.

To date, circles of this nature have been observed clearly in several areas with ancient continental crust at depth: the western United States, northernmost Mexico, the Appalachians, Alaska and the Yukon. Their existence is also strongly indicated in Madagascar and Corsica. No other areas have been examined. The circles are visible on displays which were produced from commercially available raised plastic relief maps

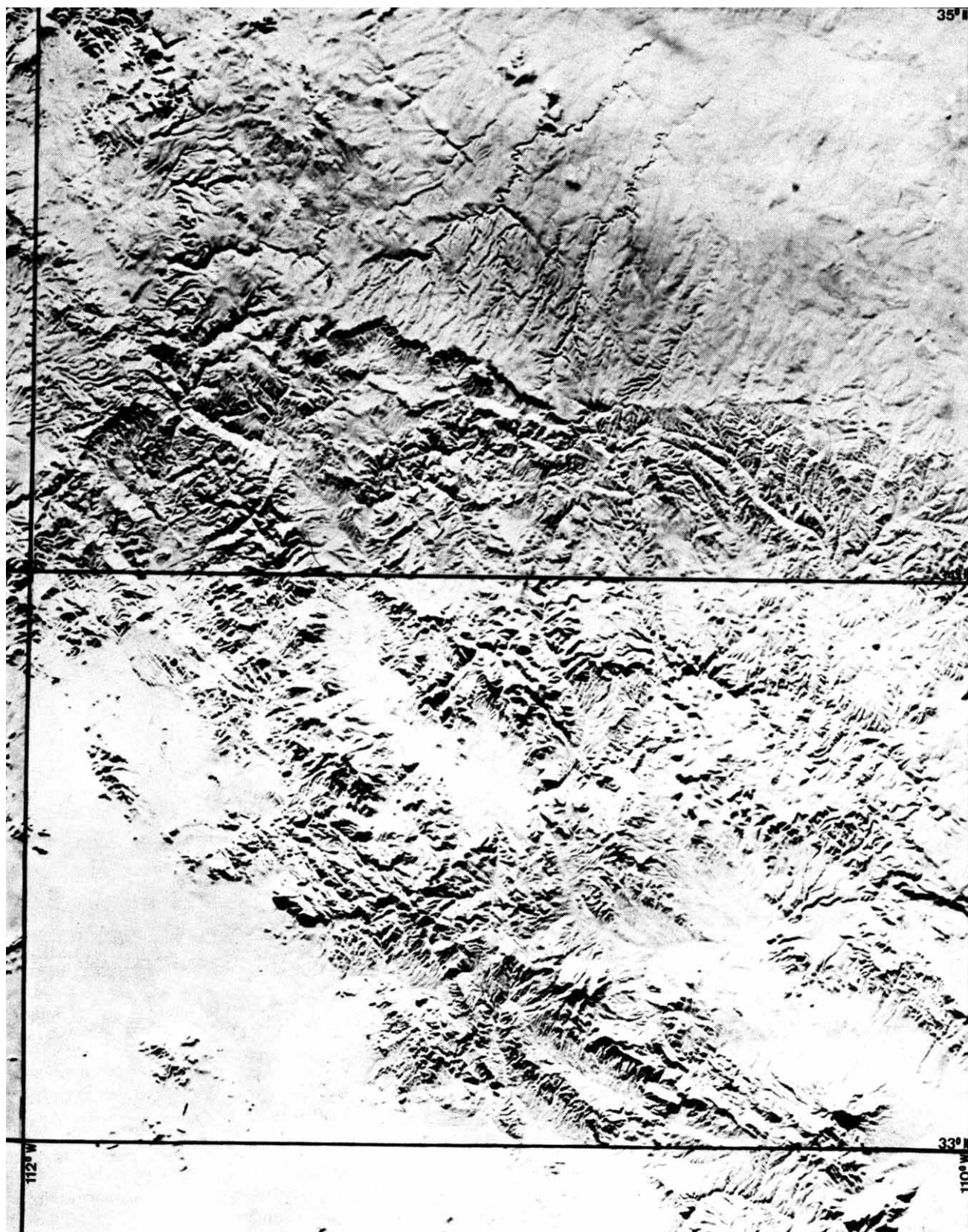


Fig. 1 Portion of Arizona at a scale of 1 : 1,000,000 produced by the method outlined in the text. The original relief maps were sprayed white and had twofold vertical exaggeration. The test area is bounded by thick latitude and longitude lines and the extra areas to the west and south are included to facilitate viewing of some of the circles.

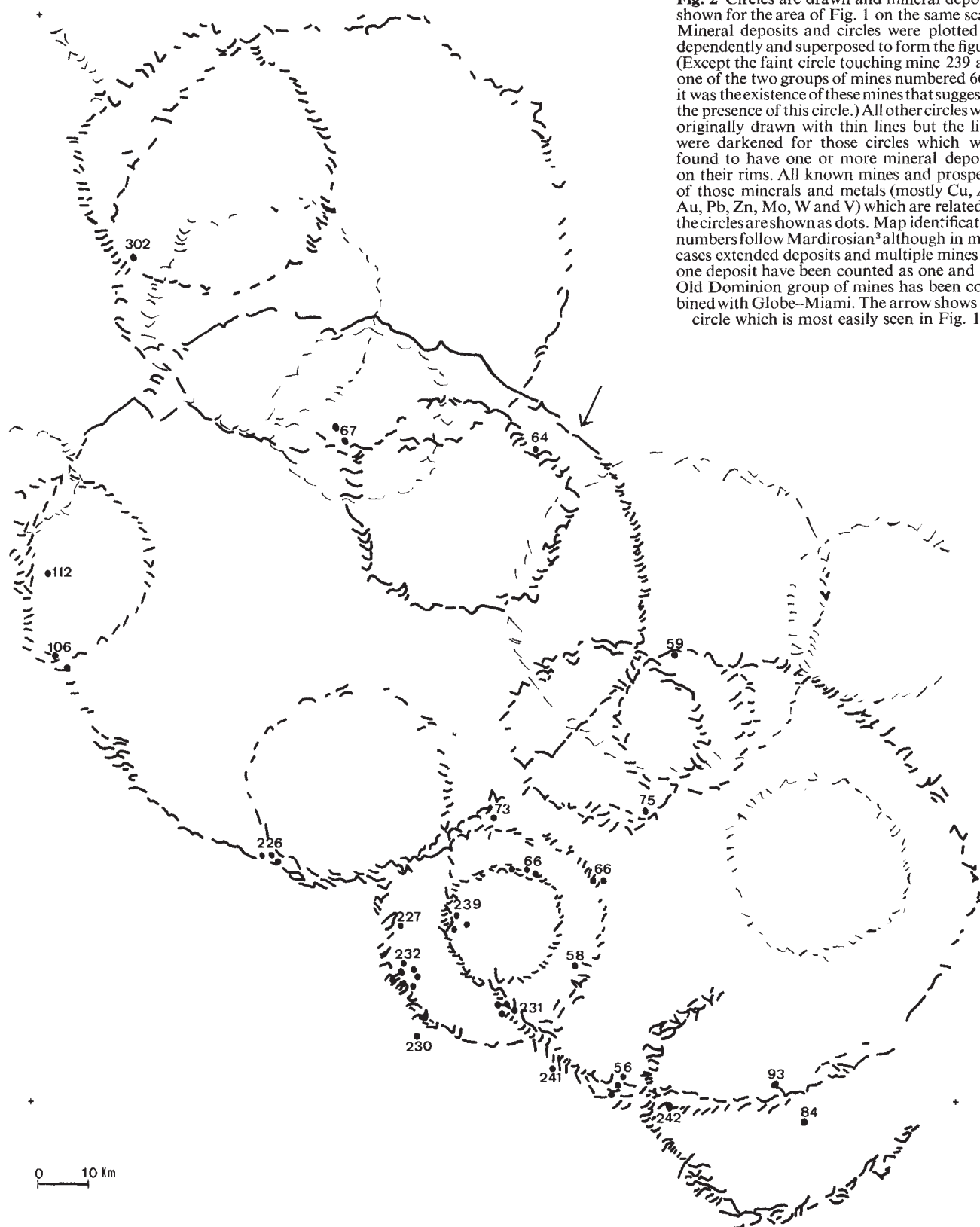


Fig. 2 Circles are drawn and mineral deposits shown for the area of Fig. 1 on the same scale. Mineral deposits and circles were plotted independently and superposed to form the figure. (Except the faint circle touching mine 239 and one of the two groups of mines numbered 66—it was the existence of these mines that suggested the presence of this circle.) All other circles were originally drawn with thin lines but the lines were darkened for those circles which were found to have one or more mineral deposits on their rims. All known mines and prospects of those minerals and metals (mostly Cu, Ag, Au, Pb, Zn, Mo, W and V) which are related to the circles are shown as dots. Map identification numbers follow Mardirosian³ although in most cases extended deposits and multiple mines on one deposit have been counted as one and the Old Dominion group of mines has been combined with Globe-Miami. The arrow shows the circle which is most easily seen in Fig. 1.

Table 1 Lode occurrences in the Arizona test area of those metals and minerals whose emplacement is controlled by the circles

District, mine, area or deposit	Principal commodities
Banner (Xmas, Troy, Dripping Springs, Hayden) (56)*	Cu, Pb, Zn, Ag, Au, Mo
Bobtail, Samsel (58)	W
Canyon Creek (59)	Turquoise
Ellison (64)	Cu, Pb, Zn, Ag
Globe (Globe-Miami); and the Old Dominion group of mines† (66)	Cu, Ag, Au, Pb, Zn, Mn, Fe, V, turquoise, chrysocolla
Green Valley (Payson), 2 deposits (67)	Au, Pb, Zn
Pinto Creek (Pinto Valley) (73)	Pb, Cu
Richmond Basin (MacMorris, McMillen) (75)	Ag
Aravaipa (84)	Pb, Zn, Cu, Ag, Au, Mo
Stanley (Stanley Butte, Quartzite Mountain) (93)	Cu, Pb, andradite garnet
Cave Creek (Gold Cliff, Golden Reef), 2 deposits (106)	Au, Cu, Ag, Pb, Mo, V, W
Magazine, Greys Gulch Mine (112)	Cu, Ag, Au
Goldfields (226)	Au, Ag
Hewitt Canyon (227)	Serpentinised limestone
Martinez Canyon (230)	Pb, Ag
Mineral Creek (Ray, Kelvin) (231)	Cu, Au, Pb, Zn, Mo
Mineral Hill (232)	Au, Pb, Zn
Pioneer (Magma, Silver King, Superior) (239)	Cu, Ag, Au, Pb, Zn, Mn
Riverside (Kearney, Rare Metals) (241)	Au, Ag, Cu, Pb, Mo
Saddle Mountain (242)	Pb, Ag
Squaw Peak (302)	Cu, Mo

The data shown are taken from Mardirosian³.

*The figures in parentheses are reference numbers taken from ref. 3 and are also used in Fig. 2.

†The Old Dominion mines have been added and are grouped with Globe-Miami.

with a 1 : 250,000 scale and a two- or threefold vertical exaggeration. These were illuminated at a low angle by a parallel light source and photographed individually. Photographs of adjacent maps were joined to give regional coverage^{1,2}. Figure 1 is an example of such a display and covers a part of Arizona.

The area shown in Fig. 1, extending from 33° to 35°N and from 110° to 112°W, covers several mining districts and was selected as a test area in which to examine the relationship between circles and mineralisation. Known mineralisation in this area includes 24 mappable occurrences or deposits of copper, silver, gold (other than alluvial), molybdenum, vanadium, tungsten, zinc, lead (other than lead occurring with alluvial gold), iron (where associated with other metals), manganese (where associated with other metals), serpentinised limestone, and turquoise. Most of the deposits are polymetallic and four are major porphyry deposits (Globe-Miami-Old Dominion, Magma-Silver King-Superior, Ray-Kelvin, Christmas-Troy-Dripping Springs-Hayden). Table 1 gives additional details. Nineteen circles were located within or partly within this test area and forty measurements of the apparent rim widths of these circles gave an average value of 2.1 km.

Using this rim width, it was found that twelve of the nineteen circles had at least one known mineral deposit along their rims and that all 24 mineral occurrences or deposits of the types listed above fell 'on' or 'very close' to a circular rim. These two terms correspond to the two highest ratings in the five point semi-quantitative scoring system used. Most of the 24 deposits were in the 9% of the total area which was classified as 'on' a rim (see Fig. 2).

The circles are not easy to see, even on the original photographs, and it is possible that a few of the faintest circles have no objective reality. There is, however, a positive correlation between degree of visibility and mineralisation. The five most visible circles have 16 out of the 24 mines or deposits associated with them. The five next most visible circles have 11 mines or deposits on or near them, 7 of which are also on or near one of the most visible circles. The nine least visible circles have only 3

deposits on them, all of which are also on or near a more readily visible circle. The degrees of visibility were assigned by J. B. Southard of the Massachusetts Institute of Technology without reference to the presence of mineralisation. One additional faint circle with mineralisation was found after the assignment of the degrees of visibility, and numerous other faint or doubtful circles were noticed subsequently. The portion of the test area covered by rims is increased from 9% to 42% if all of these are accepted.

Occurrences in the Arizona test area of asbestos, barite and iron (when not associated with other metals) also seem to bear some relationship to the circles. There was no correlation between the circular rims and the locations of deposits of uranium, manganese (in monometallic ores), perlite, placer gold, mercury, gypsum, sand and gravel, sandstone, limestone, peridot, quartz crystal, amethyst, jasper and onyx marble. No other occurrences of economic minerals or stones are reported from the area with the exception of fossil fuels³. A brief survey of circular rims found in the Butte, Montana area also showed a strong correlation with mineralisation, so the striking results from Arizona are not unique.

The correlation of mineralisation with the circles stands on its own and does not depend on the acceptance of any particular theory as to how the circles came into existence.

Individual circles exhibit different degrees of visibility. In general, a circle that is readily visible along one portion of its circumference will be readily visible over the whole circumference, and a faint circle will be uniformly faint. The great majority of circles are indeed indistinct, and there is a real problem in reproducing the same circles in different viewings of the photomaps. Larger circles are easier to see than the smaller circles and a positive correlation between circle size and presence of mineralisation is suggested. This may have an upper limit, however. In this study no major mines were noted on circles larger than 150 km in diameter. Much larger circles may also exist. A circular feature of generally similar appearance and a 2,200-km diameter seems to encircle the southern end of Africa, passing through the watershed area of central Angola, the Limpopo Valley region, the offshore edges of the Mozambique and Agulhas Plateaus, various seamounts and the Walvis Ridge.

Intersections of two circles seem to have a somewhat enhanced chance of being mineralised and thus the identification of the fainter circles may prove important after all. In a few cases, the most notable of which is the Bingham copper mine in Utah, mineralisation was found at the intersection of a circle and an obvious linear feature.

The centres of the circles seem to fall in a northwesterly trending pattern in Arizona and in a northeasterly trending pattern in the Appalachians, in both cases forming elongated clusters parallel or nearly parallel to the regional geological trends. Circles are much more abundant in mountainous areas than in the adjacent plains in many regions.

The positions of many towns, roads, railroads, rivers, lakes and reservoirs seem to be related to circles and this is probably a natural consequence of the fact that the circles are defined by topographic criteria.

There is little or no correlation between circles seen on the displays used in this study with circles seen on ERTS photography.

A final observation offers an essentially independent verification of the features discussed here: when transparent overlays with circles plotted on them are reproduced at the correct scales and placed over geological maps of the corresponding areas, the various colours representing different rock types fall into new patterns which, although previously unperceived, either on the maps or in the field, are real and require interpretation. Figure 2 can be enlarged to make a suitable overlay for the Geological Map of Arizona¹.

The circles can be interpreted as having been formed by the impacts of meteorites during the most recent major bombardment of the Earth. At the time of this bombardment, which

by analogy with the moon, was a discrete episode approximately 4,000 Myr ago, the Earth already had a brittle crust. Estimates of the thickness of that crust vary greatly. The Earth today is brittle down to a depth of at least 2.5–4 km; below this it is ductile. Impacts can cause major fracturing through to the brittle/ductile transition depth and the resulting scars may never heal completely due to a combination of several mechanisms. For instance, as the impact site is eroded away, the ductile subbasement will rise and become brittle with decreasing pressure, and the circular scars will be propagated downwards to the new brittle–ductile boundary. Similarly, movements in the fractured zones can be generated upwards to bequeath the circular patterns to overlying sediments or igneous sequences^{5,6}. A history of this nature seems capable of accounting for the preservation of some of the circular scars throughout geological time. The present-day fracture patterns may thus be inheritors or descendants of the original impact craters rather than direct remnants of the craters themselves. The circles differ from previously described astroblemes and the usual criteria of shock metamorphism have not been used for their recognition.

The process of sequential inheritance, by which the geological structures remain after the rocks in which they were formed are gone, is proposed as the mechanism which permits 4,000-Myr-old circular structures to exist in profusion on a planetary surface which holds virtually no rocks thought to be of such antiquity. The geological structures remain after the rocks in which they are formed are gone—an analogy to the smile of the 'Cheshire Cat' comes to mind. Alternative hypotheses avoid this problem but give rise to others.

(1) If the circles are interpreted as artefacts of the map maker, the correlation with mineralisation is unexplainable.

(2) Although some of the circles may be modified calderas (there is evidence suggesting that some Tertiary and Quaternary calderas may have inherited and taken over pre-existing circular sites), the circular scars have a systematically higher degree of circularity than is typical of recognised calderas, and caldera complexes and their metamorphosed equivalents would have to be very much more widespread than present geological maps indicate.

(3) If the circles have been generated from below at various times as, say, super ring-dikes, or their extrusive counterparts, why are their rocks not generally recognised as such and mapped appropriately? In addition, the apparent absence of super ring-dikes in fresher states of preservation would seem to go against the principle of uniformity. The super ring-dike explanation is also difficult to reconcile with the observation that there are few circles smaller than 7 km in diameter and none under 5 km. On the other hand, the small scars caused by impacts which did not break through to the brittle–ductile boundary would have healed over completely in time. Partially healed astroblemes might exhibit features which are, nevertheless, similar in appearance to those of large ring dikes.

Difficult questions remain even if the impact scar origin for the circles is accepted. For instance, the trends which parallel the regional geology in Arizona and the Appalachians have no clear interpretation, nor does the observation that the circles are found in greater numbers in the mountains than in the surrounding plains. Two useful speculations concerning these observations are: first, that the presence of the circular structures permitted or even caused the mountains to form where they did, or, alternatively, that the forces that formed the mountains also reopened the old scars and, second, that the limited sensitivity of the map making or photographic techniques used hinders the identification of circles in regions of low relief.

Geological activity after the formation of the circular features might have prevented their complete healing over and thus: (1) permitted the intrusions of the igneous bodies which are partly responsible for the topography of the rims, (2) permitted the introduction of mineralising fluids from below, (3) permitted the introduction of descending and oxidising surface waters from above, and (4) maintained the circular patterns on the Earth's surface.

Two main characteristics of 'porphyry' copper deposits are association with an intrusive body and, in almost all cases, supergene enrichment of the ore zone. A third characteristic is dependence on intense fracturing whose nature has not been satisfactorily explained previously^{7,8} despite the observation by Godwin in 1973 that "some breccias and adjacent rocks in the vicinity of porphyry deposits . . . have features that seem indicative of shock metamorphism"⁹. This study has shown that a fourth characteristic of porphyry deposits in the Arizona test area and elsewhere is that they are confined to circular rims. All these characteristics can be explained as consequences of an impact-scar site of formation.

If the circles are formed by impacts, they will be surface expressions of forms which, due to geometrical constraints, are generally cylindrical or perhaps somewhat conical. These are most probably essential elements of the "deeply penetrating fundamental 'plumbing' systems" which Jerome and Cook¹⁰ and others believe exert a controlling influence on the concentration of metals. For the case of the porphyry type deposits, it is evident that the circular patterns and their extensions in depth did influence the movements of various fluids, hot, cold, rising, descending, viscous, non-viscous, recent and ancient, and this suggests why other types of deposits which are formed partly through the action of thermal waters (gold-bearing quartz veins, for instance) should be preferentially located on the circular rims. A logical extension of this concept and some fragmentary evidence indicate that the scar structures also affect the distribution of other fluids, groundwater, gas and oil.

I thank Renée Lecomte for preparing the manuscript, Sylvie Salade for plotting the circles, Ruth L. Saul and Janice A. Glaholm for editing, and Professors John B. Southard (Massachusetts Institute of Technology) and Gerald J. Wasserburg (California Institute of Technology) for useful comments.

JOHN M. SAUL

ORYX,
3 rue Bourdaloue,
75009 Paris, France

Received 20 January; accepted 9 November 1977.

1. Gay, S. P., Jr *Pervasive Orthogonal Fracturing in Earth's Continental Crust*, 105–106 (American Stereo Map Co., Salt Lake City, 1973).
2. Lindsay, J. B. *Geomorphology of Arizona*, 1 : 1,000,000 (Salt Lake City).
3. Mardirosian, C. A. *Map of Mining Districts and Mineral Deposits of Arizona*, 1 : 1,000,000 (Mineral Research Co., Albuquerque, 1973).
4. Wilson, E. D., Moore, R. T. & Cooper, J. R. *Geologic Map of Arizona 1 : 500,000* (Arizona Bureau of Mines and US Geological Survey, 1969).
5. Titley, S. R. & Hicks, C. L. (eds) *Geology of the Porphyry Copper Deposits—Southwestern North America*, 7 (University of Arizona Press, Tucson, 1966).
6. Kutina, J. *Proc. 1st Int. Conf. Basement Tectonics*, 567 (Utah Geological Association, Salt Lake City, 1976).
7. Emmons, W. H., *Trans. Am. Inst. Min. Met. Eng.* 75, 797–815 (1927).
8. Grew, P. C. *Trans. Am. geophys. Union* 57, 4 (abs.) (1976).
9. Godwin, C. I. *Proc. Geol. Ass. Can.* 25, 9–12 (1973).
10. Jerome, S. E. & Cook, D. R. *Bull. Nev. Bur. Mines* 69, (1967).

Palaeolithic–Neolithic seed remains at Franchthi Cave, Greece

FRANCHTHI Cave is located in a limestone headland on the south-west coast of the southern Argolid, Greece. In excavations since 1967 (performed under the direction of Prof. Thomas W. Jacobsen, Indiana University) an almost continuous sequence from the Upper Palaeolithic ~20,000 BC to the Final Neolithic, around 3000 BC, has been uncovered¹. After six excavation seasons, eight trenches have been opened up inside the cave, as well as eight trenches outside the mouth of the cave and along the beach about 50 m below. Owing to the proximity of the deposits to the modern shore line and to the soil surface, no carbonised material has been recovered from these beach deposits. This is unfortunate because the earliest Neolithic on the site has been found in this area. Concentrated water sieving of the trenches inside the cave has recovered large quantities of carbonised botanical material. Most important among these remains has been the discovery of wild oats and wild barley in levels dated to about 10,000 BC, earlier than any botanical remains as yet found in Greece or the Aegean area.

Table 1 Dimensions of *Hordeum spontaneum*

Tell Mureybit grains*	Length	Breadth	Thickness	L:B	T:B
Average	5.44	1.89	1.26	290	67
Minimum	3.8	1.5	1.0	252	61
Maximum	6.7	2.1	1.6	372	76
Franchthi Cave grains					
Average	5.69	2.80	1.75	205	63
Minimum	4.2	2.1	1.0	164	47
Maximum	7.5	3.6	2.1	250	90

*After van Zeist and Casparie⁵.

Inside the cave, the two largest and deepest trenches, F/A and H/H-1 quadrants A and B, have been systematically water sieved for the recovery of botanical and other environmental material. Because of the fresh water springs running into Koilada Bay below the mouth of the cave, it was possible to use nearly fresh water rather than salt water for this sieving process. A dam was built and a large hose inserted at the mouth of one of the springs. Water from the spring was then pumped up into two 1 m² storage tanks installed up the slope from the dock on which the smaller sieving tanks were placed. The earth was dumped slowly into these smaller tanks which contained a nylon net mesh of ~ 1.5 mm. The water flowing into the bottom of the tank separated the soil from the rest of the material. The carbonised remains, as well as small snail shells and bones were floated off into another nylon net which was then set on specially built racks indoors to dry.

To date this flint from the south half of F/A and one quadrant of H/H-1 (A) has been sorted, giving a nearly complete sequence from the Upper Palaeolithic (P2233-21,480 ± 350 yr bp; ~ 20,000 bc (all ¹⁴C dates are calibrated with a 5,568 yr half life) through the 'Mesolithic' (a tentative designation for the material from levels dating between 8300 bc to 6000 bc) and Neolithic to the Final Neolithic (P1660-5,261 ± 64 yr bp; ~ 3000 bc). A preliminary identification has been given to most of the botanical remains from this sequence, but measurements have not yet been made on most of the seeds and positive identification for many of the species must wait for further research.

The lowest levels dating from around 20,000 bc up to 11,000 bc contain only species of the Boraginaceae family, namely *Lithospermum arvense* L., *Alkanna* cf. *orientalis* (L.) Boiss., and *Anchusa* sp. A small percentage of these have been burnt, but the majority are in a 'fresh' state. The vast numbers of these species, as many as 2,000 *Alkanna* sp. in one unit, are difficult to explain. There are several possibilities, such as the use of the plant for bedding or fuel, or the use of the root as a dye. Perhaps the seeds were not brought into the cave by man but are the result of rodent or bird activity. There are many bird bones in these levels; this may support the suggestion that these stony seeds were used as 'crop stones' by the birds (S. Payne, personal communication). Further study of the faunal and cultural material from the Upper Palaeolithic may help to solve this problem.

In deposits dated to 11,000 bc lentils were recovered. These are an extremely small species (about 2 mm diameter, but their condition and the present lack of comparative material, will not allow a positive identification to the species level. According to Zohary, both *Lens nigricans* (Bieb.) Godr. (*Ervum nigricans* Bieb.) and *L. ervoides* (Brign.) Grande (*L. lentacula* (Schreb.) Alef.) are small-seeded lentils which are typical Mediterranean elements today. *L. orientalis* (Boiss.) Hand-Mazz. has also been reported from Greece² and is more closely

related to the cultivated *L. culinaris* Medik (*L. esculenta* Moench) than the other two species³. These early lentils at Franchthi may be one of the more common Mediterranean species and are unrelated to the cultivated lentils which appear in the Neolithic deposits on this site. The possibility of their being *L. orientalis*, however, which were collected through the Upper Palaeolithic and Mesolithic and developed into the cultivated variety in the Neolithic must also be considered.

Also occurring in the Palaeolithic deposits at about the same time as the lentils are a species of vetch (*Vicia* sp.). These are small round seeds (2.5–3 mm diameter) which, again, have not been identified to the species level as the hilum is missing from all those otherwise complete enough to measure. The positive identification of the vetches from the cave will be left until it is possible to compare the samples with an adequate modern collection.

At the same time that lentils and vetches begin to appear in the Upper Palaeolithic levels, pistachio (*Pistacia* sp.) and almonds (*Prunus amygdalus* Batsch.) appear. These two species occur together throughout the entire sequence to the Final Neolithic. The *Pistacia* sp. has previously been identified as *P. atlantica* Desf⁴, but the closer examination of the seeds of *P. terebinthus* L. and *P. lentiscus* L. may indicate their presence. Complete measurements of the Franchthi material are needed before any definite species can be assigned. The asymmetry of the Franchthi nutlets suggests that these may be closer to the *lentiscus* than the *atlantica* form.

The almonds have been identified as *Prunus amygdalus* Batsch., but as there is only one whole almond in the entire sequence, it is difficult to be certain of the identification. The whole specimen is 16.5 mm long and 11.8 mm wide. The surface is slightly pitted and grooved and the keel is not prominent.

In levels dated to around 10,500 bc (P1827-12,543 ± 176 yr bp) two grains of oats (*Avena* sp.) and one of barley (*Hordeum* cf. *spontaneum* C. Koch.) have been identified. Neither species occurs again, however, until about 9000 bc in trench H/H-1 A (Palaeolithic) and 7000 bc in trench F/A (Mesolithic), from which points they become major components of the botanical assemblage through the Mesolithic.

There is a problem with the identification of the barley grains in that they are broken and so badly preserved that it is impossible to obtain complete measurements. They have been tentatively identified as *H. cf. spontaneum* on the basis of their similarity to the illustrations of the wild barley reported from the site of Tell Mureybit in Northern Syria⁵, and with modern *H. spontaneum* in the comparative seed collection at the University of Southampton. The grains from Franchthi, however, are larger than those from Mureybit as seen in a comparison of the dimensions (Table 1). Of those grains well enough preserved to measure, it is possible to describe them in much the same terms as used for the Mureybit grain. The dorsal side is longitudinally straight or slightly curved due to puffing during carbonisation. The ventral side is straight or barely convex. In every case one or both ends of the grain is broken off, but they are generally broader at the upper end, tapering toward the lower (radicle) end.

The positive identification of these grains is a crucial factor in the question of early domestication at Franchthi. If the Mesolithic barley is, in fact, a domesticated form it would be the earliest found so far in Greece. Should it prove to be a wild species, however, it raises the question as to whether it is the progenitor of the cultivated barley found in the Neolithic levels at Franchthi. In either case the material will prove to be of great importance in the study of the origins of agriculture in this area.

The oats (*Avena* sp.) will probably not be identified further as there are no lemma bases preserved in the deposits. The grains are very small, the average length being 4.2 mm for the 12 grains so far measured (Table 2). They are spindle shaped, tapering at both ends to a point. On a few of the better preserved grains the hairs are still visible. The number of different species of wild oats in Greece, and the possibility of these being

Table 2 Dimensions of *Avena* sp. from Franchthi Cave

	Length	Breadth	Thickness	L:B	T:B
Average	4.20	1.35	1.13	314	85
Minimum	3.3	1.1	.9	250	64
Maximum	5.5	1.7	1.6	363	145



Fig. 1 Botanical remains found in Franchthi Cave.

secondary grains make a definite identification impossible.

Around 8000 BC ($P2231-10,260 \pm 110$ yr bp) in trench F/A the Boraginaceae decrease substantially down to only a few seeds in each unit, and the *Pistacia* sp. increase greatly, becoming the predominant species in the Mesolithic. At one point, however, dated to about 7000 or 6500 BC, there may have been a redeposition of earlier soils as the soil matrix itself is of the reddish, clayey Palaeolithic type and there are vast quantities of *Lithospermum arvense* in these units. At the same time there is a decrease in all other botanical finds in this trench. Further study of the stratigraphy in this area is necessary before we can determine exactly what has happened at this point in the cave's occupation.

Also in the Mesolithic, beginning around 7300 BC ($P2230-9,280 \pm 110$ yr bp), wild pear (*Pyrus amygdaliformis* Vill.) begins to appear. Both whole carbonised fruits and the seeds have been preserved in these very rich deposits. In addition to the lentils and vetches, peas (*Pisum* sp.) have been identified from two Mesolithic levels⁴. As yet, no additional finds of this legume have been recovered from these trenches. There are large legumes from these Mesolithic levels which resemble more closely a *Lathyrus* sp., but positive identification must await further study.

The next substantial change in the sequence comes at about 6000 BC ($P2095-7,981 \pm 105$ yr b.p.) with the absence of oats and the first appearance of *Triticum dicoccum* Schubl., cultivated emmer wheat. The change is sudden and complete in trench F/A, occurring through a depth of about 15 cm of earth, and suggests that there may have been a stratigraphic break or a lack of occupation, at least in this part of the cave at this time. *Hordeum distichum* L., cultivated two-row hulled barley, appears shortly after the *Triticum*, both species begin to occur in steadily increasing numbers in levels dated to around

5000 BC. There is also a change in the lentil at this point and those now occurring are the larger, cultivated variety *Lens culinaris* Medik. The vetch also seems to have changed, and the species represented has been tentatively identified as *Vicia ervilia* (L.) Willd. (= *Eryum ervilia* L.). The *Pisum* sp. which had been present in the Mesolithic has not been found in the Neolithic deposits so far examined. Pistachio and almond still occur and very small numbers of Boraginaceae are present in scattered units up to the Final Neolithic.

There is no major change in the species represented throughout the Neolithic with the exception of the appearance of *Triticum monococcum* L. at about 4500 BC in the upper Middle Neolithic levels. This species continues to occur in very small numbers in the late and Final Neolithic as well. Several wild grape pips have been found in the late and Final Neolithic levels and have been attributed to the species *Vitis sylvestris* L.

The extensive water sieving operation (Fig. 1) at Franchthi Cave has therefore, produced, a sequence of botanical material from the Upper Palaeolithic (~ 20,000 BC) to the Final Neolithic (~ 3000 BC). In the Upper Palaeolithic wild species of lentils, vetches, pistachios, almonds, oats and barley were collected. These species remain the primary botanical resources recovered from the cave through the Mesolithic to the Neolithic. In addition, wild pear and peas have been identified from the Mesolithic levels.

There seems to be an abrupt change in these resources around 6000 BC when the wild oats and barley and the wild pear disappear from the botanical material so far examined and the cultivated varieties of emmer wheat, hulled two-row barley, and lentils are found. Cultivated einkorn wheat is present in small amounts from the later Middle Neolithic levels, and wild grape first appears in the Late Neolithic.

The great importance of the Franchthi material is that it

helps to fill a void in our knowledge of pre-Neolithic subsistence in Greece. No other site in this area of the Eastern Mediterranean has yielded botanical material dated to the periods before agriculture had become established. Not only can the Franchthi material answer some of the questions concerning the gathering economy preceding agriculture, but it may also be of primary importance in establishing whether or not agriculture was incipient or introduced to Greece. It has long been felt that the knowledge of cultivation had been brought to Greece from the Near Eastern 'centres'. With the appearance of both earlier, wild forms and later, cultivated forms of certain species at Franchthi Cave, there must be a rethinking of this problem of the origins of agriculture.

Although this is only a preliminary report of the Franchthi botanical material it is hoped that further work will confirm some of the tentative statements and that it will help us to understand the crucial period of subsistence change from plant collecting to plant cultivation.

These excavations were performed under the direction of Professor Thomas W. Jacobsen, Indiana University, as part of the Argolid Exploration Project cosponsored by the University of Pennsylvania and Indiana University¹.

JULIE HANSEN

JANE M. RENFREW

Department of Archaeology, University of Southampton,
Southampton, UK

Received 7 October; accepted 24 November 1977.

1. Jacobsen, T. W. *Hesperia* 38, 343–381 (1968); 42, 45–48 (1973); 42, 253–283 (1973).
2. Davis, P. H. (ed.) *Flora of Turkey and the East Aegean Islands* 3, 327 (1970).
3. Zohary, D. *Econ. Bot.* 26, 236–372 (1971).
4. Renfrew, J. M. (op. cit.) in *Hesperia* 42, 45–48 (1973).
5. van Zeist, W. & Casparie, W. A. *Acta Bot. Neerland* 17, 44–53 (1968).

Lipid sacs as a buoyancy adaptation in an Antarctic fish

THE Notothenioidei, the dominant perciform suborder of Antarctic fishes, is a predominantly benthic group of about 75 species¹. They lack a swim bladder, a feature consistent with their mode of life on the bottom of the sea¹. Therefore few of them have been able to exploit the enormous biomass of planktonic crustaceans (krill) indigenous to the mid-waters of the Antarctic Ocean^{2,3}. One notable exception, however, is the nototheniid *Pleuragramma antarcticum* Boulenger which has invaded this highly productive habitat. Unlike other members of this family, *Pleuragramma* spends its life swimming in the water^{2,4}, at times beneath a cover of sea ice. Although *Pleuragramma* lacks a swim bladder, it has been observed to remain nearly motionless in the water when placed in an aquarium. Furthermore, its slow swimming speed and general body morphology (Fig. 1) appear to preclude the use of forward motion and associated hydrodynamic lift as a mechanism for maintaining position in the water column. In -1.8°C seawater, specimens of *Pleuragramma* weighed between 0.5 and 1% of their weight in air indicating they were almost neutrally buoyant. We report here that this reduction in density results from the accumulation of lipid in intermuscular and subcutaneous sacs, and from a reduction in skeletal ossification.

Pleuragramma were collected by lowering a 2-m² net to a depth of 475 m through a 3-m² hole in the ice at McMurdo Sound, Antarctica ($77^{\circ}54'S$, $166^{\circ}40'E$). The net was retrieved at the rate of 60 m min⁻¹.

When freshly caught specimens of *Pleuragramma* were transilluminated, translucent sacs (0.5–3.0 mm in diameter) containing lipid were seen between the muscle masses at the bases of the dorsal and anal fins. Gross staining of formol-calcium-fixed specimens with oil red O (ref. 5), a stain for neutral lipids, confirmed the presence of lipid and also revealed the existence of many smaller (0.2–1.5 mm) subcutaneous sacs (Fig. 1)

on the sides of the body. Subcutaneous sacs were also seen in the midventral line between the pelvic girdle and the anterior margin of the anal fin. In cross section the intermuscular sacs were medial to the dorsal and ventral-most subdivisions of the epaxial and hypaxial muscle masses (Fig. 2). In this position they were adjacent to the median septa, pterygiophores and the associated erector and depressor muscles of the fins.

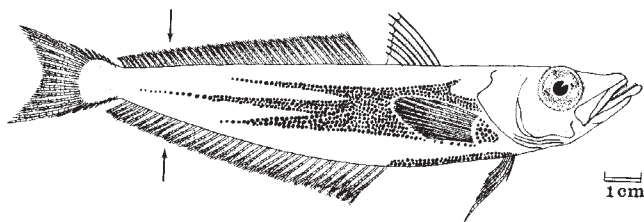


Fig. 1 *Pleuragramma antarcticum* showing distribution of subcutaneous lipid sacs (large black dots) in an adult specimen. Redrawn from Boulenger¹⁵. Arrows indicate the plane of Fig. 2.

Examination of histological sections stained with haematoxylin and eosin or picro-ponceau indicated that the lipid was enclosed in delicate connective tissue sacs, true adipose tissue being rare in fishes⁶.

For analysis, 1 ml of lipid was withdrawn from the intermuscular and subcutaneous sacs of a 30-g specimen. Thin-layer chromatography revealed that the lipid was entirely triglyceride. Extraction and analysis of the total lipid content of the trunk musculature showed that 39% of the dry muscle mass was lipid, with 75% of that being triglyceride (personal communication from P. W. Williams). The methyl esters of the triglyceride fatty acids were prepared⁷ and quantified by gas-liquid chromatography on a column of 10% SP 2340 resin on Chromasorb WAW. The fatty acid composition of the total triglyceride isolated from the muscle was similar to that of the triglyceride present in the sacs.

The density of triglyceride is approximately 0.93 at 0°C . This large amount of extracellular lipid must significantly reduce the density of *Pleuragramma* and would appear to be involved in buoyancy regulation, as is the case with other mid-water fishes which lack swim bladders⁸. The localisation of the lipid in preserved specimens is consistent with this hypothesis; the sacs were largest and most numerous near the centre of gravity of the fish.

Accumulation of lipid was not the only mechanism observed to play a part in density reduction in *Pleuragramma*. Specimens

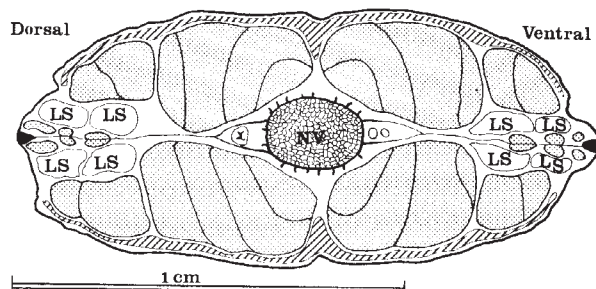


Fig. 2 Cross section of *P. antarcticum* showing location of intermuscular lipid sacs. Hatching, red muscle fibres of lateralis superficialis; stippling, epaxial and hypaxial muscles, erector and depressor muscles of fins; solid black, bone (pterygiophores, vertebral collar and trabeculae); LS, intermuscular lipid sacs; NV, notochordal vesicles (do not contain lipid). Bony tissue in the vertebrae is confined to a thin peripheral collar buttressed by longitudinal trabeculae running the length of the column. These appear as small, lateral projections in this cross section. Notochordal material fills the entire vertebral body and is continuous throughout the extent of the column. The dorsal and ventral median septa connect the tips of the neural and haemal spines to the pterygiophores.

cleared and stained with alizarin red S⁹ had reduced and poorly ossified skeletons compared with other Antarctic nototheniids. For example, the vertebrae were not amphicelous, but merely a thin collar of bone surrounding, and barely constricting, a large core of gelatinous material (Fig. 2), probably representing a persistent notochord. Furthermore, neural and haemal arches and spines were reduced, and ribs were very small. In addition, the scales were thin and weakly mineralised. These observations were further supported by ashing the bones for 10 h at 450 °C. The ashed skeleton of *Pleuragramma* constituted only 0.3–0.5% of the total body weight compared with 0.9% for a closely related but less actively swimming species, *Trematomus borchgrevinkii*. In most fishes the ash content of the skeleton is about 2% of the total body weight¹⁰.

It has been suggested before that lipid deposits in liver^{11,12}, bone^{13,14}, integument¹⁴ and muscle¹⁴ play a role in buoyancy regulation. In such cases, however, the precise role of lipids is obscured by the fact that these organs are complicated structures with primary functions unrelated to buoyancy control. In *Pleuragramma* the utility of the lipid sacs in buoyancy regulation can be more easily assessed for two reasons. First, because they are distinct and isolated structures, the only other function attributable to the sacs would be lipid storage before metabolic utilisation. This, however, seems unlikely because the lipid is not in fat cells. Second, there seemed to be an obvious correlation between the appearance of lipid sacs and the degree of skeletal ossification. That is, in small specimens (less than 44 mm) with very slight ossification of the skeleton, subcutaneous and intermuscular lipid sacs were not present. In somewhat larger (45–60 mm) specimens exhibiting some ossification, a moderate number of sacs were present. In large adults showing considerable skeletal ossification, the sacs were extensively developed. The lipid accumulating with increase in size of the fish should serve to offset increasing density resulting from ossification.

Thus it seems that both intermuscular and subcutaneous lipid accumulation and a reduction in skeletal ossification have evolved as specialisations to reduce density in *Pleuragramma*, a pelagic Antarctic fish without a swim bladder. These adaptations enable this fish to maintain the neutral buoyancy necessary for swimming and feeding in the mid-waters of the Antarctic Ocean—an ecological niche under-utilised by fishes.

We thank Ed Osada for helping net the *Pleuragramma*. This research was supported by the USNSF.

ARTHUR L. DeVRIES

Department of Physiology and Biophysics,
University of Illinois,
Urbana, Illinois 61801

JOSEPH T. EASTMAN

Division of Biology and Medicine,
Brown University,
Providence, Rhode Island 02912

Received 23 June; accepted 17 November 1977.

1. Marshall, N. B. in *Biologie Antarctique* (eds Carrick, R., Holdgate, M. & Prévost J.) 273–278 (Hermann, Paris, 1964).
2. Andriashev, A. P. in *Biogeography and Ecology in Antarctica* (ed van Oye, P. & van Miegheem, J.) 491–550 (Junk, The Hague, 1965).
3. Nikolsky, G. V. *The Ecology of Fishes* (Academic, London, 1963).
4. DeWitt, H. H. in *Antarctic Ecology I* (ed. Holdgate, M. W.) 305–314 (Academic, London, 1970).
5. Humason, G. L. *Animal Tissue Techniques* third edn (Freeman, San Francisco, 1972).
6. Tashima, L. & Cahill, G. F. Jr in *Handbook of Physiology, Section 5: Adipose Tissue* (eds Renold, A. E. & Cahill, G. F., Jr) 55–58 (American Physiological Society, Washington DC, 1965).
7. Morrison, W. R. & Smith, L. M. *J. Lipid Res.* 5, 600–608 (1964).
8. Marshall, N. B. *The Life of Fishes* (World Publishing Co., Cleveland and New York, 1966).
9. Taylor, W. R. *Proc. U.S. natn. Mus.* 122, 1–17 (1967).
10. Lagler, K. F., Bardach, J. E. & Miller, R. R. *Ichthyology* (Wiley, New York, 1962).
11. Bone, Q. & Roberts, B. L. *J. mar. Biol. Ass. U.K.* 49, 913–937 (1969).
12. Malins, D. C. & Barone, A. *Science* 167, 79–80 (1970).
13. Lee, R. F., Phleger, C. F. & Horn, M. H. *Comp. Biochem. Physiol.* 50B, 13–16 (1975).
14. Bone, Q. *Copeia* 78–87 (1972).
15. Boulenger, G. A. in *Report on the Collections of Natural History Made in the Antarctic Regions During the Voyage of the "Southern Cross"* 174–189 (British Museum (Natural History), London, 1902).

Equalisation of prey numbers by migratory shorebirds

STUDIES of the effects of predators on prey diversity have usually focused on patterns of predation that do not by themselves increase diversity. For example, size-selective fish decrease diversity by driving large bodied zooplankton to local extinction¹. Size-selective predators can maintain a diversity of species only if large bodied prey hold some advantage over smaller prey, such as the difference in feeding efficiency hypothesised by Brooks and Dodson (ref. 1 and see ref. 2 for review). The diversifying effect of nonselective predation on competitors for space in the rocky intertidal zone³ depends on the presence of differences among competitors in time of settlement during the year. In contrast, selective removal of common prey tends to reverse any numerical advantage gained by one species over another. Thus it is of interest to see whether numerically selective predators increase diversity by equalising prey numbers. The results presented here show that general predators in fact repeatedly equalise the relative abundance of prey species and hence tend to stabilise prey communities.

Migratory shorebirds are flexible predators^{4, 5} that attain considerable numbers in estuaries along the Atlantic coast of North America. Daily counts of shorebirds in the Plymouth, Massachusetts estuary exceeded 5,000 birds from mid-July to mid-August in 1975 and 1976. Most birds belonged to one of four species: sanderling (*Calidrus alba*), semipalmated sandpiper (*Calidrus pusilla*), short-billed dowitcher (*Limnodromus griseus*), and black-bellied plover (*Pluvialis squatarola*).

Each of these species used several foraging techniques at Plymouth, and each showed some degree of habitat selection. Numerical selectivity was evident from examination of crops and gizzards, which generally contained the two or three commonest prey species in the area where the bird had been foraging before collection. Large and abundant invertebrates (*Clymenella torquata*, *Scoloplos robustus*, *Nereis virens*, *N. arenaceodonta*, *Tellina agilis*, *Crangon septemspinosa*, *Acanthostatorius millisi*, *Trichophoxus epistomus*) were well represented in shorebird diets. Numerical dominants of smaller size (spionid polychaetes, *Gemma gemma*, *Hydrobia (totteni)*, *Protohaustorius deichmannae*) were under-represented relative to their abundance on the flats. Both prey size and prey number seem to be important in the selection of prey by shorebirds⁶.

Invertebrate mortality was estimated by counting the organisms in 20 core samples taken at each of 13 representative sites in July, as birds began increasing in number, and comparing this with the count of organisms at the same sites in September, after most shorebirds had departed. September samples were located within a metre or two of July samples to reduce the effect of a patchy spatial distribution on estimates of invertebrate loss. The sampling procedure itself did not lower invertebrate density, based on the lack of a significant difference between two rounds of sampling at one site on a day in September. All cores were 10 cm in diameter and 10 cm deep. Cores were washed on a 0.5-mm sieve and the organisms visible by eye were counted and recorded in the field. This procedure captures at least 90% of the organisms greater than 2 mm, based on re-examination of samples in the laboratory. The effect of juvenile recruitment on estimates of invertebrate mortality was controlled by recording the size of organisms with sufficient accuracy to distinguish 0-yr class from older animals.

Enclosure experiments in 1975 and 1976 showed that losses between July and September were due to predation. Each enclosure was 1 m on a side, 10 cm high, and made of 14-mm mesh wire. The number of adult invertebrates inside the cages did not change significantly between July and September ($P > 0.50$), compared with the significant declines immediately outside the cages during the same period ($P < 0.05$). Wire cages also exclude horseshoe crabs (*Limulus polyphemus*) and

flatfish (*Pseudopleuronectes americanus*). Flatfish are largely absent from Plymouth Harbour during July and August⁷, when they move offshore into deeper, cooler water⁸. Usage of the study sites by horseshoe crabs did not correspond to invertebrate losses, based on a Spearman rank correlation across the 18 out of 26 0.5-hectare sites where crabs occurred. Mortality at sites used heavily by shorebirds exceeded mortality at less used sites.

Mortality in both years depended on the rank abundance of prey (Table 1). Higher mortality occurred in the initially more abundant prey at a site even though the ranking of individual species in July differed from site to site. Of the numerically dominant macrofauna only *Gemma gemma* did not show substantial losses when abundant. This small (< 4 mm), thick-shelled clam is the commonest species at many sites but it occurs at a low frequency in the gizzards of shorebirds, which suggest that it is relatively unpalatable.

Table 1 Mortality of intertidal invertebrates fed on by migratory shorebirds in Plymouth and Kingston Harbors in 1976

Species rank within sites in July	1	2	3-5	> 5
Average mortality (1-final/initial)	0.84	0.78	0.67	0.36
Average initial density (organisms per 10 cores)	90.9	26.0	26.3	14.1
No. of 0.5 hectare areas	26	26	26	26
No. of different species	6	11	21	32

Mortality is the difference between counts at a site in mid-July and mid-September, divided by the initial count and averaged over sites. All cores were 10 cm in diameter and 10 cm deep. Variation in mortality with frequency was analysed by grouping species according to their rank in July in 10 cores from an area of one-half of a hectare. The change in density within each rank is statistically significant ($P < 0.05$). Changes in mortality between adjacent rankings are all significant at $P < 0.01$, based on a test of the equality of proportions.

Frequency dependent losses resulted in an equalisation of prey numbers during the summer. Equitability ($J' = H'/\ln(s)$, where s is the number of species and H' is the Shannon-Weaver index of diversity⁹), increased significantly from July to September of 1976 ($P < 0.001$), based on a nonparametric Wilcoxon signed ranks test¹⁰. This increase remains significant when the negative association between equitability and collection size has been removed by regression. A similar increase in equitability occurred in 1975. The coefficient of variation¹⁰, another measure of dispersion, decreased significantly in both years.

Shorebirds reduce variation in the relative abundance of their prey by selective removal of numerically dominant species. This would tend to halt the continued expansion of dominant species, since few adult prey escape predation and since these prey reproduce only during the spring and summer. However, differences in palatability make it unlikely that shorebirds could by themselves stabilise an entire community of intertidal invertebrates.

This work was supported in part by contract 14-16-0008-687 from the US Fish and Wildlife Service to the Manomet Bird Observatory. I thank Brian Harrington and the Observatory staff for their help.

DAVID SCHNEIDER

Manomet Bird Observatory, Manomet,
Massachusetts 02345 and
Ecology and Evolution, State University of New York,
Stony Brook, New York 11794

Received 16 September; accepted 28 November 1977.

1. Brooks, J. L., & Dodson, S. I. *Science* **150**, 28 (1965).
2. Hall, D. J., Threlkeld, S. T., Burns, C. W. & Crowley, P. H. A. *Rev. Ecol. Syst.* **7**, 177 (1976).
3. Paine, R. T. *Am. Nat.* **100**, 65 (1966).
4. Baker, M. C. *Ecology* **55**, 162 (1975).
5. Goss-Custard, J. C. *J. Anim. Ecol.* **46**, 21 (1977).
6. Griffiths, D. *Ecology* **56**, 1209 (1975).

7. Iwanowicz, H. R., Anderson, R. W. & Ketschke, B. A. *Monograph 17 Massachusetts Department of Natural Resources* (1974).
8. McCracken, F. D. *J. Fish Res. Bd Can.* **20**, 551 (1963).
9. Peilou, E. C. *An Introduction to Mathematical Ecology* (Wiley, New York, 1969).
10. Sokal, R. R. & Rohlf, F. J. *Biometry* (Freeman, San Francisco, 1969).

Separation of seed development from monocarpic senescence in soybeans

MONOCARPY, a single flowering phase followed by senescence and death, is a widespread phenomenon in seed plants, particularly field crops, whose large monocultures put on a dramatic display during their senescence phase¹⁻⁴. In many, but apparently not all monocarpic species, defruiting or deflowering can prevent or at least delay monocarpic senescence and death¹⁻⁴. For some species such as soybeans, the killing influence has been traced to the developing seeds⁵. The conspicuous correlation between the accumulation of nutrients in the developing fruit and the senescence (or apparent 'exhaustion') of the leaves (Fig. 1) has led to the suggestion that developing fruits cause senescence by diverting or withdrawing needed nutrients or hormones from the leaves and other vegetative parts¹⁻⁴. Here we explore the relationship between nutrient (or hormone) withdrawal or diversion and foliar senescence. Our data indicate that the development (accumulation of dry matter) of the seeds is separable from the senescence response in soybeans, and therefore the seeds may function as more than sinks.

Soybeans (*Glycine max* (L.) Merrill) cv. Anoka were grown in soil as described by Lindoo and Noodén⁶. During the first 4 weeks, the plants were kept in a greenhouse with supplementary light at night and thereafter in environmental control chambers under short (10-h) days. Foliar senescence (yellowing) and fruit development (growth and colour) were measured quantitatively by rapid, non-destructive, visual procedures, which have been checked against the more traditional (but destructive) methods for measuring senescence and fruit development described in detail elsewhere⁶. Fruit development is expressed as fruit maturity index (FMI), which reflects the average state of fruit development based on numerical values 1 (least mature) to 5 (most mature).

Our earlier studies⁵ on soybeans showed the influence of the developing seeds on foliar senescence moves in a restricted pattern within the plant, thereby distinguishing it from the flower-inducing hormone, the movement of which is not as restricted and differs in other ways⁷.

Single axis soybean plants with pods only on the bottom half and foliage only on the top half show greatly reduced foliar senescence compared with unmodified plants and that which occurs is primarily in the leaves nearest to the pods (Fig. 1a). The seed yield of the pod-bearing sections of these plants relative to unmodified plants is not reduced and is, if anything, slightly increased (unpublished data). Subsequent depodding of these 'half-and-half' plants at various stages of fruit development causes an additional reduction in senescence depending on when the pods are removed. Early (up to short day 58) removal of the pods from the bottom half stops foliar senescence, and removal at short day 71 or later does not substantially delay that senescence which does occur.

Figure 1b shows the time course for seed growth (dry weight per plant) and the total seed weight at the depodding. It is important to note that the dry weight of the seeds is nearly maximum (90%) at short day 66 but comparing Fig. 1a with b, it can be seen that depodding, nevertheless, causes an almost total suppression of foliar senescence. These data suggest not only that seed growth can occur without producing foliar senescence, but the senescence is induced mainly during the final phase of seed development when nutrient accumulation is probably complete. The

spatial separation of leaves and pods seems to be important here, for normal plants with leaves subtending the pods do not show this clear distinction between seed growth and senescence when depodded at various stages of seed growth⁶.

When plants are modified to a single leaf and a single pod cluster separated by two nodes, the foliar senescence pattern differs strikingly, depending which is on top, the leaves or the pods (Fig. 2). When the pods are on top, the leaf frequently senesces, whereas it does not when the

pods are below the leaf. This is consistent with our earlier observation that the senescence signal tends to move downward if at all⁵. Neither delayed fruit development nor reduced seed yield can account for the lack of senescence when the single leaf is on top, the time course of fruit maturation in these modified plants is only 4–10 d behind the unmodified control (podded) plants; the yield per node is actually greater in the plants with a single pod cluster than in the unmodified control (Table 1); and the yield per node is the same whether the pod-bearing node is above or below the leaf (Table 1). Complete defoliation at FMI 1.5 reduces the seed yield in these single pod clusters by 80%; most of that seed growth which does occur apparently occurs before defoliation. The single remaining leaf seems to supply required nutrients for the growing seeds and to be subject to whatever nutrient deprivation is created by the seeds.

The experiments shown in Figs 1 and 2 show in different ways that seed development (and therefore nutrient deprivation) can occur without monocarpic senescence.

The reproductive sink size (seed yield) can be controlled by removing all except a defined number of pods from plants at FMI 1.9 (Fig. 3). In this variety and in the conditions used, seed yield (dry weight) was nearly directly proportional to the number of pods per plant; there was no pronounced tendency for seed size to increase in compensation for a reduced number of seeds. The seed nitrogen per plant is likewise proportional to the number of pods, yet foliar senescence is maximum when the number of pods is only about 40% of the maximum. A direct proportionality between sink size and rate of senescence would be expected if monocarpic senescence were caused by nutrient deprivation; in particular, the saturation of senescence induction at very low pod doses is inconsistent with the nutrient deprivation theories.

The possible explanations for the correlative influence of the developing seeds on foliar senescence fall into four

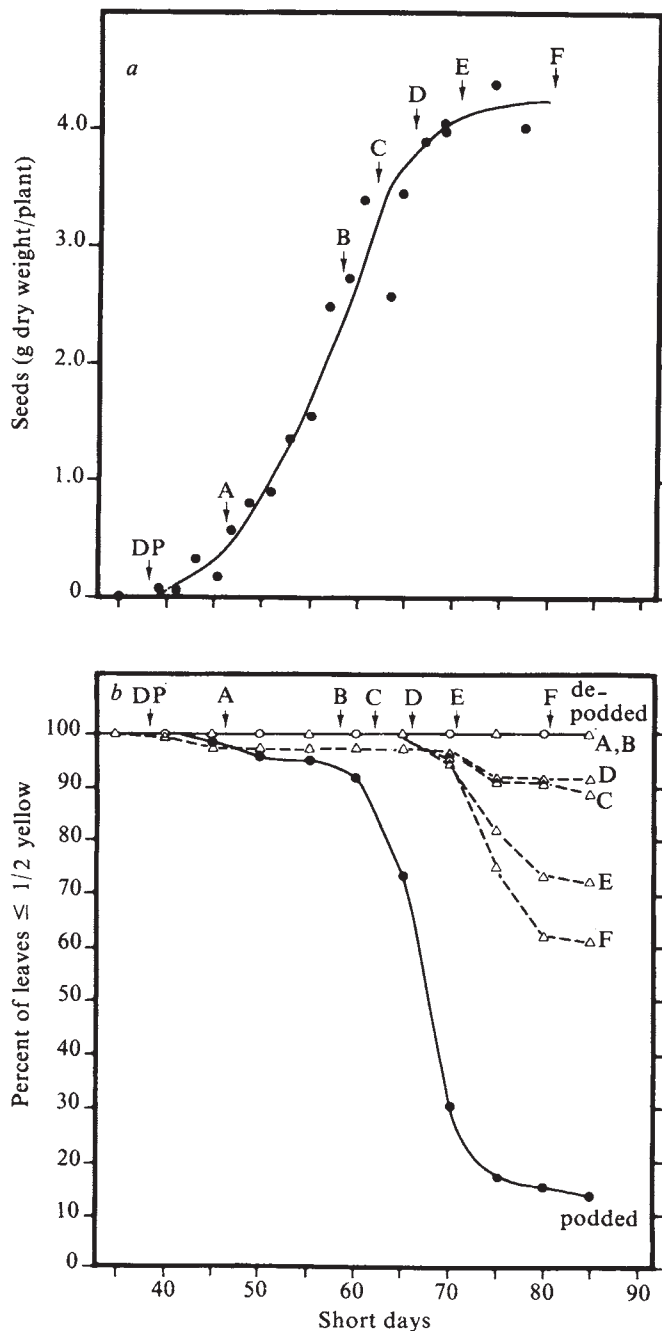


Fig. 1 *a*, Time course of seed growth in surgically modified, single-axis soybean plants (leaves only on top half, pods only on bottom half) and the effect of removing the remaining pods at various stages (indicated by A, B, C, and so on) of seed growth. Surgical modification at FMI 1.4 with subsequent removal of pods as they reached the start of pod fill, stage 2. Measurement of senescence, fruit development and FMI are explained in ref. 6. A single node between the top and bottom halves was both depodded and defoliated. *b*, Time course of foliar senescence in single-axis soybeans with leaves only on the top half and pods only on the bottom half. Same plants as in (*a*). Plants designated 'depodded' were completely depodded at FMI 1.4, whereas the 'podded' plants were not modified at all.

Fig. 2 Time course of foliar senescence and fruit maturation for single-axis plants pruned to one leaf and one pod cluster (separated by three nodes) compared with unmodified plants and depodded plants with all leaves intact. Depodded and defoliated at FMI 1.5 with subsequent removal of pods as they reached stage 2.

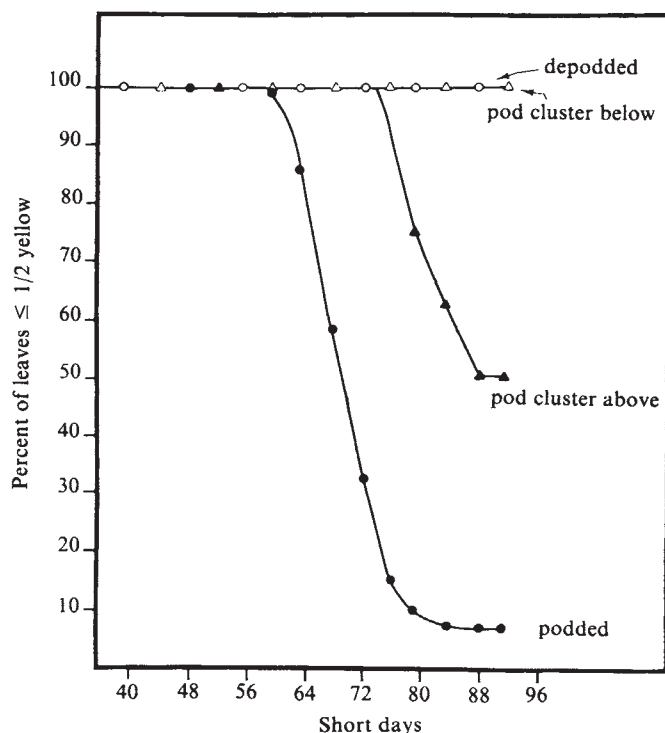


Table 1 Seed yields on single-axis soybeans with only one leaf and one pod cluster* or without modification

	Seeds (g dry weight/node) Unmodified plant	Modified plant*
One pod cluster three nodes above a single leaf (or a pod cluster on a comparable node of an unmodified plant)	1.6	1.5
One pod cluster three nodes below a single leaf (or a pod cluster on a comparable node of an unmodified plant)	0.9	1.9

*Same plants as in Fig. 2. Unmodified plants yield about 8.3 g of dry seeds each.

general categories, which have some variations and combinations⁴ but can now be evaluated in the light of the data for soybeans. The first, and least precisely defined, holds that the plants simply have a limited lifespan which runs out about the time the seeds develop. One variant of this hypothesis invokes a limited lifespan for the meristems, the loss of which causes death. Although it may be true that the shoot apical meristems degenerate during flower development thereby limiting vegetative growth (this point needs study in soybeans), depodding of soybeans at an early stage can prolong the life of these plants way beyond normal, possibly indefinitely, even though they undergo little or no vegetative growth. Thus the senescence and death we have observed in conditions favourable for soybean growth is coupled to seed production and not time or meristem degeneration.

The second hypothesis has the developing fruit diverting nutrients (for example, minerals from the roots, photosynthetic assimilate from the leaves) or hormones (cytokinins from the roots) away from the leaves thereby causing their death. Diversion of photosynthetic assimilate from the leaves seems an unlikely cause, for mature or senescent leaves seem to produce more than they need^{10,11}. The diversion hypothesis, particularly the diversion of substances from the roots, is difficult to reconcile with the limited

movement of the senescence 'signal' from the seeds⁵. If such a diversion were important, then why does the single pod cluster cause senescence of the single feeder leaf only when the feeder leaf is below the pods and not when above? One would expect the pods to divert substances from the roots more effectively when they are below the leaf. Furthermore, the idea that foliar senescence is triggered by a reduction in the supply of substances produced by the roots, is also countered by the observation that foliar senescence occurs in these plants where the leaf- or pod-to-root ratio is greatly reduced. In addition, the delayed depodding of single-axis plants with pods only on the bottom half and leaves only on the top half suggests that senescence induction by the seeds may occur after most of the seed growth and its concomitant nutrient demands are complete. Note that in order to prolong the life of the plants and to maintain their productivity; expert gardeners take care to remove the fruits from certain varieties of cucumbers and beans after they fill out but before their colour changes (for example, after drain or diversion is completed but before monocarpic senescence is initiated). Finally, the sink size (pod dose or seed yield) does not parallel foliar senescence; the senescence response is saturated at a level far below the maximum level of dry weight and nitrogen accumulation in the seeds.

The third hypothesis involves nutrient drain; the developing fruit are supposed to withdraw nutrients from the leaf to the point of killing them. Indeed, there probably is some exodus of certain nutrients (particularly nitrogen) from senescing leaves⁴. The arguments against diversion, however, apply equally well to drain.

The fourth explanation gives the developing fruit or seeds a more active role: production of a substance(s) which travels out to the leaves to cause senescence and ultimately the death of the whole plant.

Although drain and diversion may be involved in the monocarpic senescence of soybeans, it seems unlikely that developing seeds exert such remarkable correlative influences simply by functioning as passive sinks, just as shoot apices do not exert apical dominance solely through nutrient diversion. Such a hormone(s) will, of course, remain hypothetical unless actually isolated and shown to travel from the source (the seeds) to the target (leaves and other parts).

Since monocarpic senescence must have evolved independently in a wide range of groups, it may well be caused by different mechanisms in other species. Indeed, the flowers of male spinach induce monocarpic senescence, even though they do not produce seed². Deflowering retards this senescence, and neither nutrient drain nor diversion seem to be causal. On the other hand, it has been reported that defoliation of cocklebur does not even delay the senescence and death of the whole plant¹².

This study was supported in part by a research grant (416-15-79) from the USDA Cooperative State Research Service under P.L. 98-106. We thank Rohbe Crotz for help in growing the plants during their first 8 weeks and Gary Kahanak for help in completing one set of experiments.

L. D. NOODEN
D. C. RUPP
B. D. DERMAN*

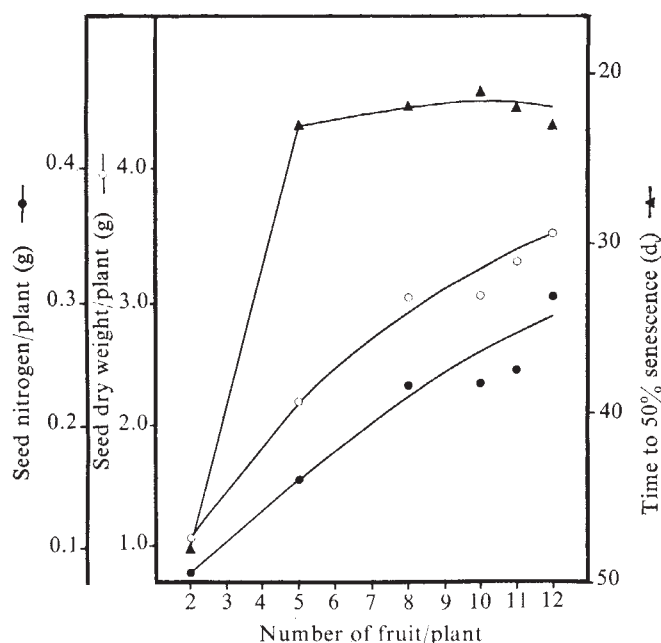
Botany Department,
University of Michigan,
Ann Arbor, Michigan 48109

Received 27 June; accepted 28 November 1977.

*Present address: Department of General Botany, University of Istanbul, Suleymaniye, Istanbul, Turkey.

1. Molisch, H., *Die Lebensdauer der Pflanze*. (G. Fischer, Jena, 1928); (transl. Fulling, E. H.) (Science Press, Lancaster, Pennsylvania, 1938).
2. Leopold, A. C., Niedergang-Kamien, E. & Janick, J. *Plant Physiol.* 34, 570-573 (1959).
3. Wangemann, E. in *Handbuch der Pflanzenphysiologie* 15 (2) (ed. Ruhland, W.) 1037-1057 (Springer, Berlin, 1965).

Fig. 3 Relationship between sink size (pod number) and rate of foliar senescence. The pods were pruned to the designated number at FM1 1.9, and thereafter extras were removed as they reached stage 2. The average number of pods on unmodified plants was 14. Total nitrogen was measured by digesting samples at 280 °C in concentrated H₂SO₄ containing CuSeO₄ catalyst followed by dilution with H₂O and addition of Nessler's reagent for colorimetric determination^{8,9}.



4. Noodén, L. D. & Leopold, A. C., in *Plant Hormones and Related Compounds* (eds Letham, D. S., Higgins, T. J. & Goodwin, P. B.) (Elsevier, Amsterdam, in the press).
5. Lindoo, S. J. & Noodén, L. D. *Plant Physiol.* **59**, 1136–1140 (1977).
6. Lindoo, S. J. & Noodén, L. D. *Bot. Gaz. Chicago* **137**, 218–223 (1976).
7. Borthwick, H. A. & Parker, M. W. *Bot. Gaz. Chicago* **100**, 374–387 (1938).
8. Dawson, R. M., Elliot, D. C., Elliot, M. H. & Jones, K. M. *Data for Biochemical Research*, 2nd edn (Clarendon, Oxford, 1969).
9. Middleton, K. R. *J. appl. Chem.* **10**, 281–286 (1960).
10. Thrower, S. L. *Aust. J. biol. Sci.* **15**, 629–649 (1962).
11. Thrower, S. L. *Symp. Soc. exp. Biol.* **21**, 483–506 (1967).
12. Krizek, D. T., McIlrath, W. J. & Vergara, B. S. *Science* **151**, 95–96 (1966).

Regulation of RNA synthesis in early germination of isolated wheat (*Triticum aestivum* L.) embryo

THE change from dormancy to germination of seeds requires water and an appropriate temperature. Wheat embryo germination is characterised by an initial increase in fresh weight, followed by a 5-h lag¹. Although RNA synthesis is one of the earliest biologically measurable activities in wheat embryo germination^{2,3}, its mode of regulation has not been established. Reports^{3–6} have suggested that the rate of synthesis does not change during the first 6 h of germination. But no adjustment was made for the specific activities of ribonucleoside triphosphate, UTP. This would have led to a misinterpretation of incorporation data when there was either a change in the concentration of the nucleotide or a difference in the rate of phosphorylation from nucleoside to nucleotide as germination progressed. To investigate further whether there is a change in the rate of RNA synthesis in the germination phase, we have used recently-developed enzyme assays^{7–9} to measure picomole changes in purine pyrimidine ribonucleotide levels. We found a threefold increase in the rate of RNA synthesis of wheat embryos germinated for 40 min–5.5 h. We used the reaction catalysed by uridine 5'-diphosphoglucose pyrophosphorylase to determine the content of UTP. The key step in this analysis is the selective adsorption of the reaction product, UDP-¹⁴C-glucose, on to activated charcoal in the presence of 0.8 M Trizma base⁹.

In a 15-min pulse-labelling experiment, a 5.1-fold increase in the rate of ³H-uridine incorporation into RNA was observed in wheat embryos germinated for 40 min–5.5 h (Table 1, experiment 1). When incorporation data were corrected by adjusting for the specific activity of the UTP, there was a 2.6-fold increase in ³H-uridine incorporation into RNA during the germination phase. When germinating wheat embryos were pulse-labelled for 15 min longer followed by RNA isolation and purification, there was a 2.9-fold increase in 5.5 h (Table 1, experiment 2), similar to the 2.6-fold increase of experiment 1. These results show that the rate of RNA synthesis increases threefold during the period of germination between 40 min and 5.5 h. These results are in contrast to the idea that RNA synthesis does not change during the first 6 h of germination^{3–6}.

The rate of RNA synthesis in the early stage of the germination phase could be limited by either (1) RNA polymerase activity; (2) changes in the template activity of chromatin, or (3) changes in purine and pyrimidine ribonucleoside triphosphate levels. Possibilities (1) and (2) are less likely because the extraction and assay of RNA polymerase does not change for the first 6 h¹¹, and Yoshida and Sasaki¹² reported that the template activity of chromatin is essentially constant for the first 18 h of germination.

There is the following support for possibility (3). To explain the difference in RNA-synthesising capacity between embryos germinated for 40 min and those germinated for 5.5 h, the concentrations of purine and pyrimidine ribonucleoside triphosphates were measured. We found that the cellular level of ATP and GTP increased 60% and 30%, respectively. This minor increase in purine ribonucleoside triphosphates does not corroborate the threefold increase in the rate of RNA synthesis during this period. However, when the pyrimidine ribonucleotide pool was assayed, there was a 400% increase in the level of pyrimidine ribonucleotides during the period between 40 min and 5.5 h of germination (Table 2). The increase of UTP was similar to that of CTP. Thus this large increase in the substrate levels of UTP and CTP could account for

Table 1 Rates of RNA synthesis in wheat embryo during the germination phase

Experiment 1	Labelling periods	
	40–55 min	5.5–5.75 h
5% TCA-insoluble radioactivity	4,130±176 c.p.m.	20,900±377 c.p.m.
Specific activity of UTP	(1)	(5.1)
Adjusted incorporation	1,788±71 c.p.m. nmol ⁻¹	3,511±211 c.p.m. nmol ⁻¹
	(1)	(2.6)
Experiment 2	30–60 min	5.25–5.75 h
RNA fraction	20,410±1,160 c.p.m.	115,040±5,640 c.p.m.
Adjusted incorporation	(1)	(5.6)
	(1)	(2.9)

To prepare the trichloroacetic acid (TCA)—insoluble extract, isolated wheat (*Triticum aestivum* L. variety Fortuna) embryos (50 mg) were germinated in a Petri dish (60×15 cm) on four layers of 4.25-cm diameter Whatman No. 1 filter paper with 2 ml of water (containing chloramphenicol, 20 µg ml⁻¹) in the dark at 25 °C for 40 min or 5.5 h. The top layer containing the embryos was removed, blotted and placed in 0.5 ml of water containing chloramphenicol (20 µg ml⁻¹) and 50 µCi of purified 5-³H-uridine (25 Ci mmol⁻¹) for 15 min at 25 °C. The paper was blotted and the embryos were rinsed with water. Embryos were ground first with 0.5 ml of 10% TCA and then with 5 ml of 5% TCA. The homogenate was centrifuged for 10 min at 23,500g and the supernatant was kept. The pellet was rinsed three times by suspending in 5 ml of 5% TCA with a motor-driven Teflon homogeniser and centrifuging for 5 min at 23,500g. The pellet was then suspended in 1 ml of 0.1 N NaOH and counted in 10 ml of Triton X-100 toluene-based scintillation fluid. The TCA-soluble extract was prepared as before⁸. Unlabelled UTP (1 µmol) was added to the extract and concentrated down to dryness at 40 °C in a flash evaporator. A quarter of the sample was spotted on a PEI-cellulose TLC (Brinkmann) which was then developed in either 0.5 M KH₂PO₄, pH 4.5, or 0.4 M LiCl, pH 5.5. The UTP region on the PEI-cellulose was removed from the TLC sheet by treating with 1 ml of 2 M LiCl. The suspension was then mixed with 10 ml of liquid scintillation fluid (ACSTM, Amersham/Searle). The endogenous UTP was determined as before¹⁴. To prepare the RNA fraction, 50 mg of isolated wheat embryos was germinated as before and labelled for 30 min with 70 µCi of 5-³H-uridine. The embryos were ground first with 0.5 ml of high salt buffer (1 M KCl, 100 mM Tris-Cl, pH 7.6, and 5 mM EDTA). Sodium dodecyl sulphate buffer (3 ml of 1% SDS, 10 mM Tris-HCl, pH 7.6, and 5 mM EDTA) was added and the mixture was vortexed for 2 min. It was then extracted with water-saturated chloroform: phenol (1:1, v/v). The aqueous fractions were combined, mixed with two volumes of absolute ethyl alcohol and stored overnight at -20 °C. The precipitate was centrifuged at 23,500g for 20 min, rinsed three times with 5 ml of cold absolute ethyl alcohol, and then dissolved in 1 ml of 0.5% SDS buffer (0.5 M NaCl, 10 mM Tris-HCl, pH 7.6, 0.5% SDS and 1 mM EDTA). Sample of 40 µl was mixed with 10 ml of liquid scintillation fluid (ACSTM) and counted as before. The data represent an average of duplicate samples. Numbers in parentheses signify the fold increase of the 5.5-h values above those of the 40-min values.

Table 2 Cellular levels of nucleoside triphosphates in early germination of wheat embryos

Germination time	ATP	GTP pmol per mg of embryo	UTP	CTP
0	8	46	<1	<1
40 min	1,288 (1)	320 (1)	132 (1)	84 (1)
3 h	2,000 (1.6)	304 (1)	420 (3.2)	200 (2.4)
5.5 h	2,072 (1.6)	425 (1.3)	540 (4.1)	320 (3.8)

Adenine nucleotides were analysed by the method of Cheung and Marcus⁷. The key reaction is the phosphorylation of ADP by ³²P-PEP in a reaction catalysed by pyruvate kinase, with the extent of transfer of ³²P to ADP determined by adsorbing the nucleotides on to charcoal. ATP is determined as ADP after being converted quantitatively by hexokinase in the presence of excess glucose. Guanine ribonucleotides were determined according to Cheung and Marcus⁸. The principle of the assay is the conversion of GDP to ³²P-GTP utilising ³²P-PEP in the pyruvate kinase reaction. ³²P-ATP which is also formed as a major product is removed with glucose and hexokinase. CTP was determined with RNA polymerase^{13,14}. A reaction mixture in a volume of 100 µl, containing 12 mM MgCl₂, 80 mM KCl, 60 mM Tris-HCl, pH 7.6, Triton X-100 (0.02%), 0.9 U of *E. coli* RNA polymerase, 17.5 µg of calf thymus DNA, 50 nmol of GTP, 50 nmol of UTP, 2 nmol of ATP, 2.5 µCi of 8-³H-ATP (17 Ci mmol⁻¹), and 50–200 pmol of CTP was incubated for 1 h at 37 °C. The suspension was then diluted with 5 ml of cold 5% TCA and filtered through glass fibre filters (Whatman GF/C). The filters were washed with 5 ml of 5% TCA, three times, dried and counted in 10 ml of toluene based scintillation solution. Figures in parentheses are as in Table 1.

the increase in RNA-synthesising capacity in the late germination phase.

Although the template activity of chromatin is essentially constant for the first 18 h of germination and active RNA polymerase is already present in the resting embryo, the amount of ribonucleoside triphosphates present in the resting embryo is almost negligible (Table 2). On the basis of these observations, we conclude that the rapid increase in the cellular content of purine and pyrimidine ribonucleoside triphosphates triggers the initial RNA synthesis once the wheat embryo is exposed to germination conditions.

We therefore propose that there are at least two stages of RNA synthesis in the early germination of wheat embryos. In the first stage (from 0 to 40 min), there is an increase in the four ribonucleoside triphosphates. In the second stage (from 40 min to 5.5 h), only the pyrimidine nucleoside triphosphate levels increase significantly. Thus the synthesis of RNA initially depends on the presence of all four ribonucleoside triphosphates. This is followed by a stage in which the pyrimidine ribonucleoside triphosphates determine the extent of RNA-synthesising capacity.

The mode of regulation of RNA synthesis in early germination of wheat embryo could be used as a working model for other biological systems. Changes in purine or pyrimidine pools may be used to explain the data of Hentschel and Tata¹⁵. They observed a differential activation of free and template-engaged RNA polymerase I and II during the resumption of development of dormant *Artemia gastrulae*. The level of uridine triphosphate has also been shown to control the growth in ascites hepatoma cells¹⁶.

We thank Dr Abraham Marcus for wheat embryos. This research was supported by a grant from the NSF.

CHI P. CHEUNG
ROBERT J. SUHADOLNIK

Department of Biochemistry,
Temple University School of Medicine,
Philadelphia, Pennsylvania 19140

Received 28 June; accepted 30 November 1977.

- Marcus, A., Feeley, J. & Volcani, T. *Plant Physiol.* **41**, 1167–1172 (1966).
- Spiegel, A. & Marcus, A. *Nature* **256**, 228–230 (1975).
- Dobranska, M., Tomaszewski, M., Grzelczak, Z., Rejman, E. & Buchowicz, J. *Nature* **244**, 507–509 (1973).

- Spiegel, S., Obendorf, R. & Marcus, A. *Plant Physiol.* **56**, 502–507 (1975).
- Chen, D., Schultz, G. & Katchalski, E. *Nature new Biol.* **231**, 69–72 (1971).
- Chen, D. & Osborne, D. *Nature* **226**, 1157–1160 (1970).
- Cheung, C. P. & Marcus, A. *Analyt. Biochem.* **69**, 131–139 (1975).
- Cheung, C. P. & Marcus, A. *FEBS Lett.* **70**, 141–144 (1976).
- Cheung, C. P. & Suhadolnik, R. J. *Analyt. Biochem.* **83**, 52–56 (1977).
- Marcus, A. *Symp. Soc. exp. Biol.* **23**, 143–160 (1969).
- Mazus, G. & Brodnieicz-Proba, T. *Acta biochim. Pol.* **23**, 261–267 (1976).
- Yoshida, K. & Sasaki, K. *Plant Physiol.* **59**, 497–501 (1977).
- Maxson, R. E., Jr & Wu, R. S. *Eur. J. Biochem.* **62**, 551–554 (1976).
- Jameson, A. T. & Webb, D. R. *Analyt. Biochem.* **72**, 212–219 (1976).
- Hentschel, C. & Tata, J. *Devl Biol.* **57**, 293–304 (1977).
- Kepler, D. *Cancer Res.* **37**, 911–917 (1977).

Response properties of dipteran giant visual interneurons involved in control of optomotor behaviour

GIANT homolateral interneurons in the third optic ganglion of flies respond to visual stimulation with graded potentials only, lacking spike generation for signal transmission in physiological conditions¹. (The term 'homolateral' is used here in its anatomical connotation of neurones which are anatomically restricted to one side of the brain⁵.) Graded potentials were thought to be the means of signal transmission of all three classes of homolateral giant elements of this neuropile region, that is, the 'vertical cells' (V cells), the 'centrifugal horizontal cells' (CH cells) and the 'horizontal cells' (H cells): both CH and H cells are mainly sensitive to horizontal pattern movement, whereas the V cells respond mainly to vertically moving patterns^{1,2}. (The CH cells have an additional sensitivity to vertical, some V cells to horizontal pattern motion.) However, there is experimental proof that these findings might be valid only for the CH cells in which exclusively graded potentials could be recorded showing discrete excitatory and inhibitory postsynaptic potentials^{1,2}. For the V cells, however, it has been shown that spiking can be induced by application of hyperpolarising currents³. In the absence of hyperpolarising currents, graded potentials with superimposed 'local' action potentials of small amplitude were observed^{1,2}. The H cells, on the other hand were reported to generate only 'local' action potentials superimposed on to postsynaptic slow potentials¹: these findings¹ are not compatible with the previous³ results or with the results presented here which demonstrate regular action potential activity in H cells in physiological conditions.

Intracellular recordings were obtained from various sites along the axons of horizontal cells. The site of penetration was not marked but could be estimated with high accuracy by using external landmarks of the neuropile for reference—for example tracheal patterns and borders of the ganglia. After obtaining the recordings, cells were marked by iontophoretic injection of the fluorescent dye Procion yellow M4RAN so that individual cells could be identified anatomically.

Figure 1 shows two of the three types of horizontal cells, one north cell (NH) and two equatorial cells (EH) following the terminology of Pierantoni⁴. Note that even the very fine dendritic processes are completely stained as comparisons with cells stained by other methods show^{5,6}.

Figure 2 shows intracellular recordings from the three cells shown in Fig. 1. In all three cases, the stimulus consisted of a moving pattern projected on to a circular milky glass screen. This moving pattern stimulated either the contralateral or the ipsilateral eye—the H cells have a binocular sensitivity—subtending a viewing angle of approximately 70° at the fly's eye. For all three recordings shown, the direction of pattern motion was such that the H cells reacted with excitatory responses. Figure 2a shows a d.c. recording obtained from the axon close to the dendritic tree in the lobula plate (see arrow, cell EH1. The cell's response was elicited by ipsilateral stimulation with a progressively moving pattern (that is, movement from the anterior to the posterior). A similar response can be obtained by recording from the dendritic branches in the lobula plate. It is clearly seen that strong excitatory and inhibitory postsynaptic potentials are superimposed, revealed by the steep depolarising

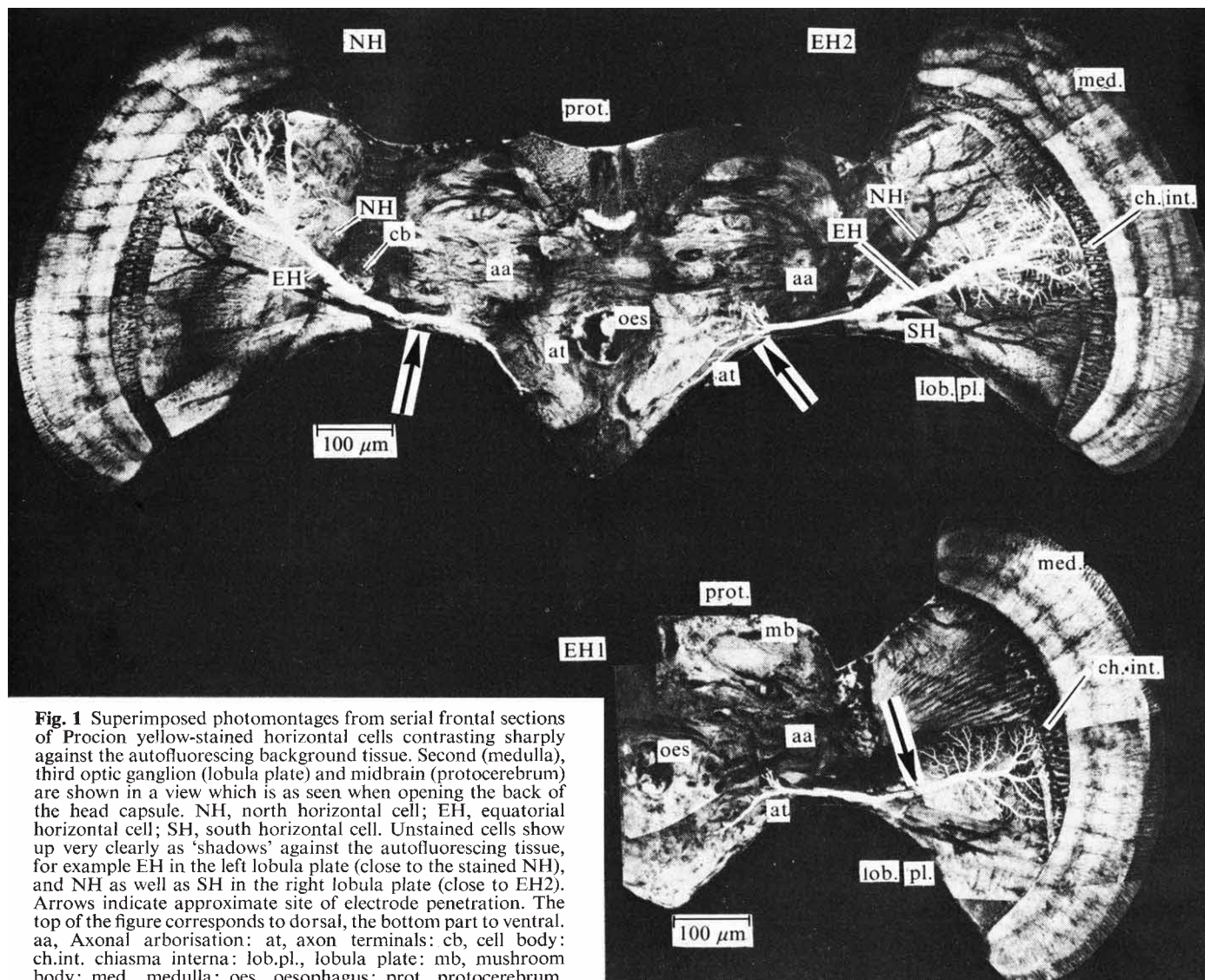


Fig. 1 Superimposed photomontages from serial frontal sections of Procion yellow-stained horizontal cells contrasting sharply against the autofluorescing background tissue. Second (medulla), third optic ganglion (lobula plate) and midbrain (protocerebrum) are shown in a view which is as seen when opening the back of the head capsule. NH, north horizontal cell; EH, equatorial horizontal cell; SH, south horizontal cell. Unstained cells show up very clearly as 'shadows' against the autofluorescing tissue, for example EH in the left lobula plate (close to the stained NH), and NH as well as SH in the right lobula plate (close to EH2). Arrows indicate approximate site of electrode penetration. The top of the figure corresponds to dorsal, the bottom part to ventral. aa, Axonal arborisation; at, axon terminals; cb, cell body; ch.int. chiasma interna; lob.pl., lobula plate; mb, mushroom body; med., medulla; oes., oesophagus; prot., protocerebrum.

and hyperpolarising potential flanks, respectively. In addition, a depolarising average membrane potential shift is observed. Figure 2b shows a recording from the axon of NH closer to the protocerebrum (in the optic peduncle connecting the lobula plate with the protocerebrum), that is, the recording was obtained close to the axonal arborisation (aa in Fig. 1) in the posterior slope. Presenting regressively moving patterns to the contralateral eye, very distinct excitatory postsynaptic potentials and rapid 'spike-like' potentials (the 'local action potentials' of Hausen¹) can be observed. For very effective stimuli, these spike-like potentials become superimposed. The ability of the axon to sum these spike-like potentials has been considered to be caused by a passively conducting membrane^{1,2}. In other words, the response shown in Fig. 2a and b can be obtained at any location along the axon. This is easily understood by calculating the space constant of such an axon which yields $\lambda = 1,000 \mu\text{m}$ (refs 1,2).

However, Fig. 2c shows a recording from a site proximal to the axonal arborisation (aa, Fig. 1). Again a regressively moving pattern was presented to the contralateral eye, using the same stimulating condition as for the experiments shown in Fig. 2b. It can be clearly seen that the cell is capable of generating regular regenerative action potentials; no slow membrane potentials accompany the action potentials. A refractory period was always detected—that is, we never observed 'superimposed spikes'. These results were not found to be cell specific, but representative for all H cells.

With high probability we can exclude the possibility that these findings (particularly the spiking activity of the axonal terminals of the H cells or the graduated potentials of the axon connecting lobula plate and protocerebrum) are injury potentials due to microelectrode penetration or to other unnatural conditions of the preparation. No changes in response properties were observed even after up to 80 min of intracellular recording time, and high frequency firing rates were not seen when cells were damaged by penetration. Furthermore, typical cell responses could be obtained from other cells even up to 2 h after obtaining recordings such as those shown in Fig. 2. However, definite proof would require the demonstration that the animal is capable of, for example, optomotor responses after penetration of such a cell.

These findings are consistent with the concept (to be fully discussed elsewhere (H.E.K., in preparation)) that (1) spikes are generated in the axon part between the axonal arborisation (aa) and the axonal terminals (at) and thus the 'local spikes'¹ reflect the spike activity of the terminal axon part since these spikes are conducted retrogradely in the passive axon membrane distal to the axonal arborisation. (2) In addition to the ipsilateral input in the dendritic fan of the lobula plate, there is a second synaptic input region, which is most likely situated at the axonal arborisation and/or the axon terminals, and which is made by contralateral elements (see also ref. 1).

Thus the giant axon connecting the dendritic fan of the lobula plate with the input region in the protocerebrum could be used

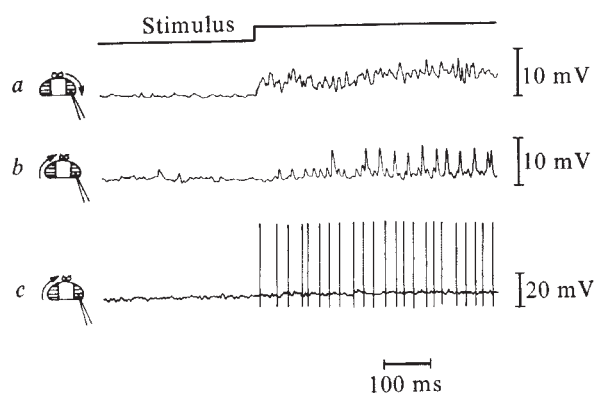


Fig. 2 Intracellular recordings obtained from the H cells shown in Fig. 1 show responses to ipsilaterally presented progressive (a) and to contralaterally presented regressive pattern motion (b,c). Pattern parameters were: spatial wavelength, 20° ; contrast, $m = 48\%$; velocity, $w = 5 \text{ deg s}^{-1}$, luminance 4 cd m^{-2} . Arrows at the schematic horizontal cross section of the head (to the left of the response traces) indicate the direction of pattern movement.

for integrating ipsilateral as well as contralateral inputs. It is not clear whether the spike activity shown by these fibres represents the output of these cells or the contralateral input, but the former seems to be the more plausible interpretation³.

This work was supported by the Deutsche Forschungsgemeinschaft. Part of the work was carried out at the University of Southern California Department of Biological Sciences, Los Angeles, and supported by grants from the NSF (BMS 74-21712) and NIH (1 ROL EY 01513-01).

HENDRIK E. ECKERT

Ruhr University Bochum,
Department of Animal Physiology,
PO Box 10 21 48,
463 Bochum-Querenburg, FRG

Received 2 August; accepted 17 November 1977.

1. Hausen, K. Z. *Naturforsch.* 31c, 629–633 (1976); Thesis, Univ. Tübingen (1976).
2. Eckert, H. & Bishop, L. G. *J. comp. Physiol.* (in the press).
3. Eckert, H. *Proc. Ger. Zool. Soc. Hamburg* 69, 84 (1976).
4. Pierantoni, R. S. *Biokybernetik* 5, 157–163 (1974).
5. Strausfeld, N. J. (ed.) in *Atlas of an Insect Brain* (Springer, Berlin, 1976); In *Neural Principles in Vision* (Springer, Berlin, 1976).
6. Eckert, H. & Boschek, C. B. *Experimental Entomology Neuroanatomical Techniques* (eds Miller, T. A. & Strausfeld, N. J.) (Springer, Berlin, 1978).

Directed outgrowth of optic fibres regenerating *in vitro*

THE way in which growing nerve fibres find their destined sites of termination is largely unknown, although there are supporting arguments and documentation for both local and long-range guidance mechanisms^{1–3}. Fibres regenerating following axotomy tend to grow in the direction of the degenerating tracts, apparently guided through channels formed by debris and/or glial cells^{4,5}. Here we present evidence that the neuritic outgrowth from goldfish retinal explants is related to the orientation of cut fibres within the explant.

We demonstrated previously that optic nerve crush 10–14 d before retinal explantation will stimulate neurite production in culture⁶ and that the neurites originate from ganglion cells⁷. On the day of explantation the retina is cut into square pieces 500 μm across, which are placed on a poly-L-lysine-coated substratum⁸. Several explants are placed in each 35-mm dish (NUNC) without regard to whether it is the vitreal or the photoreceptor surface that contacts the substratum. The ganglion cell axons in these experiments have been cut twice: once, at the time of optic nerve crush, several millimetres from the ganglion cell soma; and again, close to the soma when the retina is cut for explantation.

Examination of explant cultures under phase microscopy indicated that neuritic outgrowth began within a day of explantation. The direction of outgrowth was influenced by several factors. For example, neurites grown on poly-L-lysine-coated planar surfaces (glass or plastic) inevitably curved in a clockwise direction. Experiments designed to reveal the nature of this tendency led us to the conclusion that the directionality reflects an inherent helicity of the fibres⁹.

The clockwise directionality of neuritic growth is apparent in Fig. 1, which also indicates another property of the explants: the early outgrowth was usually restricted to one side or corner (Fig. 1a). A probable explanation for this asymmetric outgrowth was provided by explants stained with Holmes silver nitrate¹⁰. An array of parallel fibres could be seen in the optic fibre layer (Fig. 2), and the region of densest neuritic outgrowth was often aligned with this axis. Since the optic fibres in the goldfish retina, as in other vertebrates¹¹, are arranged in straight, radial lines, all of which converge at the optic disk where the fibres exit the eye as the optic nerve, it seemed probable that the parallel fibres within the explant were the ganglion cell axon tracts. Within the small area of the explant, convergence of the fibres was not apparent. Thus, it was not possible to determine which was the centripetal end of the fibre axis, so we could not establish whether the outgrowing fibres were indeed pointing at the phantom optic disk. In a brief, undocumented report 30 years ago, Vinnikov described directed outgrowth from retinal explants of several vertebrate species which he attributed to the radial pattern of optic fibres¹².

In order to establish definitively whether the neurites were aligned with the pre-existing ganglion cell fibre tracts within the explant, we performed the experiment represented schematically in Fig. 3. A strip of retina approximately 750 μm wide, extending across the full expanse of the retina from one peripheral margin to the other and including the optic disk, was prepared and then cut into thirds. Unlike our usual explants,

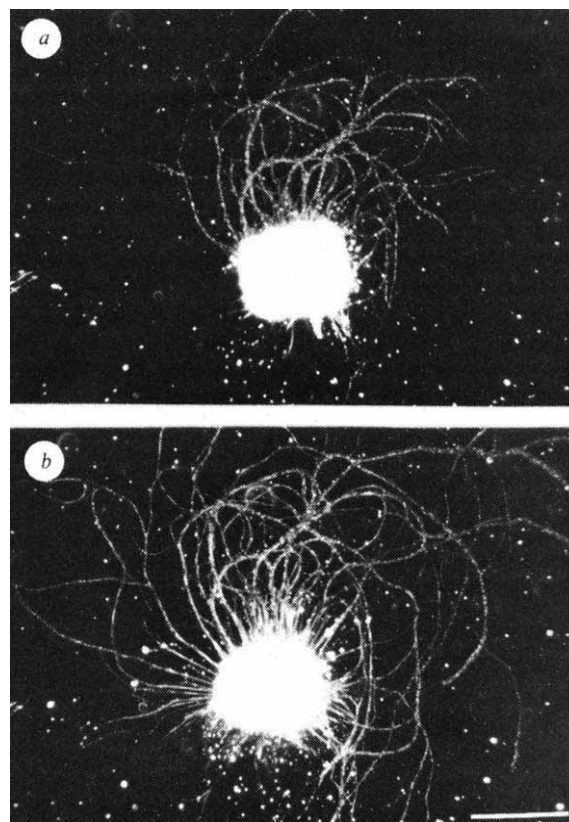


Fig. 1 Retinal explant viewed with darkfield illumination. The majority of neurites were restricted to one edge after 4 d *in vitro* (a). Note that most of them curve in a clockwise direction. After 14 d (b), neurites had grown out from the other edges as well. Scale bar, 500 μm .

the orientation of each of these pieces was known. We expected no growth from the edges corresponding to the peripheral margins of the retina and very little from the two lengthwise edges, since the latter intersected few of the 'spokes' formed by ganglion cell axons. Further, we anticipated little growth from the central explant, since its ganglion cell axons would be aimed inward. The experimental results shown in Fig. 4 indicate that the neurites grew in the predicted direction. The pattern of outgrowth shown was observed in several replications and was independent of the placement of strips with respect to one another in the dish. The observed asymmetry of early outgrowth was thus a property of each piece of retinal tissue and did not reflect interaction between explants or other extrinsic factors.

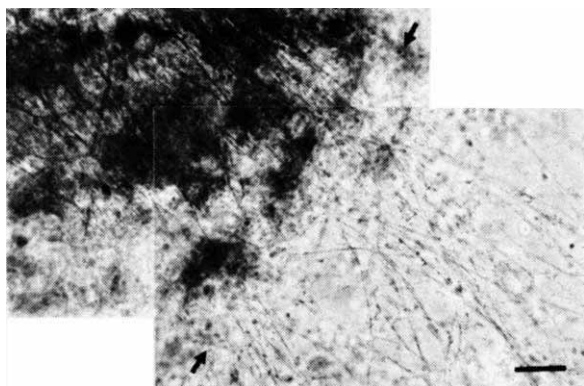


Fig. 2 In this silver-stained explant (6 d *in vitro*), the optic fibre layer can be seen at the upper left; the edge of the explant lies between the two arrows. Neurites aligned or contiguous with fibres in the explant have grown out on to the substratum at the lower right. Scale bar, 25 μ m.

After several days in culture, the asymmetry of outgrowth was diminished as fibres emerged from the entire circumference of both the small, square explants (Fig. 1b; see also previous studies^{7,9}) and from the large strips. Examination of silver-stained explants offered a possible explanation for the eventual appearance of outgrowth from edges which were initially bare. Many of the later-appearing fibres had wandered through the explant or were deflected at the tissue-substratum boundary and grew along the edge of the explant for some distance before emerging on to the dish. In contrast, the first neurites seen probably regenerated from axonal stumps whose cut ends were lined up along the edge of the explant nearest the optic disk, so that their path was unimpeded as they grew out on to the substratum.

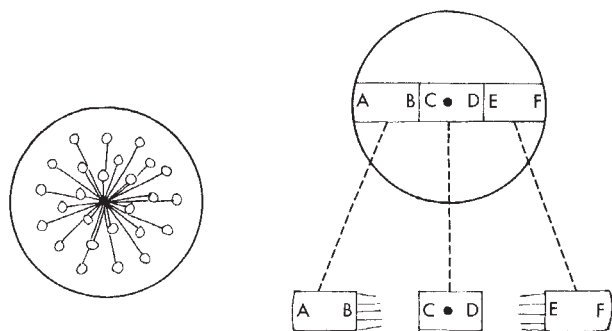


Fig. 3 The schematic diagram on the left shows the radial pattern of optic fibres in the retina. Ganglion cell somata are represented by small, open circles and their axons by line segments which converge at the optic disk (filled circle). At the upper right, the retina is shown as it was prepared for explantation. A strip of retina that included the optic disk was divided longitudinally into three approximately equal pieces and then explanted. If the neurites maintained their centripetal orientation (towards the disk), they would grow out, as shown at the bottom right, from the ends labelled B and E.

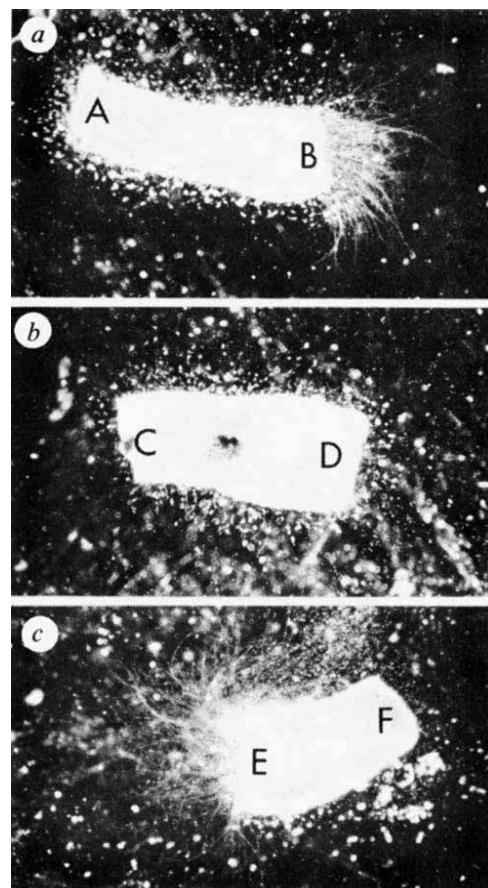


Fig. 4 The results of an experiment as described in Fig. 3 are illustrated. After 6 d *in vitro*, most of the neurites have grown out as predicted from ends B (a) and E (c). The optic disk appears as a dark spot in b. Scale bar, 500 μ m.

The explants in Fig. 4 each contain approximately 5,000 ganglion cells¹³ from which axons regenerated in a direction predictable from their pre-existing pathways. While it is well known that growing axons can be guided both *in vivo* and *in vitro*^{1,14,15}, in these retinal explants a great many fibres have been channelled in the same direction in the absence of extrinsic influences. This preparation thus seems to provide a unique opportunity to directly examine retinal fibre-fibre interactions and their proposed role in neuronal recognition^{16,17}.

We thank Dr Anne M. Heacock for useful suggestions. This study was supported by NSF grants BMS 75-03810 and NIH MH 12506-10 to B.W.A. P.R.J. was supported by NIH postdoctoral fellowship NS 05518-01.

PAMELA RAYMOND JOHNS
MYONG G. YOON*
BERNARD W. AGRANOFF

Neuroscience Laboratory,
University of Michigan,
Ann Arbor, Michigan 48109

Received 26 September; accepted 28 November 1977.

*Present address: Department of Psychology, Dalhousie University, Halifax, Nova Scotia, Canada.

1. Weiss, P. J. *exp. Zool.* **68**, 393-448 (1934).
2. Ramón y Cajal, S. *Degeneration and Regeneration of the Nervous System* 362-392 (Hafner, New York, 1959).
3. Hunt, R. K. & Jacobson, M. *Cur. Top. dev. Biol.* **8**, 203-259 (1974).
4. Murray, M. J. *comp. Neurol.* **168**, 175-196 (1976).
5. Turner, J. E. & Singer, M. J. *exp. Zool.* **190**, 249-268 (1974).
6. Landreth, G. E. & Agranoff, B. W. *Brain Res.* **118**, 299-303 (1976).
7. Johns, P. R., Heacock, A. M. & Agranoff, B. W. *Brain Res.* **140** (in the press).
8. Letourneau, P. C. *Dev. Biol.* **44**, 77-91 (1975).
9. Heacock, A. M. & Agranoff, B. W. *Science* **198**, 64-66 (1977).
10. Wolf, M. K. J. *Cell Biol.* **27**, 259-279 (1964).
11. Rodieck, R. W. *The Vertebrate Retina: Principles of Structure and Function* 477-479 (Freeman, San Francisco, 1973).
12. Vinnikov, J. A. *Nature* **158**, 377 (1946).
13. Johns, P. R. & Easter, S. S. *J. comp. Neurol.* **176**, 331-342 (1977).

14. Jacobson, M. *Developmental Neurobiology*, 136-151 (Holt, Rinehart and Winston, New York, 1970).
15. Sidman, R. L. & Wessels, N. K. *Expl Neurol.* **48**, 237-251 (1975).
16. Hope, R. A., Hammond, B. J. & Gaze, R. M. *Proc. R. Soc. B* **194**, 447-466 (1976).
17. Levinthal, F., Macagno, E. & Levinthal, C. *Cold Spring Harb. Symp. quant. Biol.* **40**, 321-331 (1976).

Neuromuscular transmission is adequate in identified abnormal dystrophic muscle fibres

It has been suggested that 'functional denervation', a failure of structurally intact neuromuscular junctions to evoke a muscle fibre action potential following nerve stimulation, plays an important part in the pathogenesis of murine muscular dystrophy^{1,2}. Other workers, however,

have been unable to confirm the existence of 'denervated' muscle fibres in dystrophic mouse muscle³. Moreover, the demonstration that transmitter release in response to nerve stimulation is normal in such muscle⁴ seems to preclude the possibility that 'denervation' may arise in an unpredictable or spasmodic way as a result of changes in transmitter store size or mobilisation during or after the repetitive discharge of the motoneurone. It has been considered possible, however, that the microelectrode techniques used in many of these studies do not always sample diseased muscle fibres, but select only a hypothetical group of 'normal' or 'healthy' muscle fibres. We have therefore applied a technique of intracellular staining^{5,6} and demonstrated that structurally abnormal muscle fibres in dystrophic muscle display normal neuromuscular transmission.

Experiments were carried out at room temperature on extensor digitorum longus (EDL) muscles removed from 3-6-months-old dystrophic (Bar Harbor 129 ReJ) mice and their clinically normal litter mates. In some muscles, muscle fibre action potentials were generated in response to indirect excitation and were recorded using standard intracellular techniques. In other muscles, endplate potentials (e.p.ps) were recorded following nerve stimulation at frequencies of 3 Hz and 30 Hz in the presence of 0.6-1.2 μ M *d*-tubocurarine. The quantum contents of the e.p.ps were estimated from the coefficient of variations of e.p.p. amplitudes⁷. After recording either action potentials or e.p.ps, a second micro-electrode filled with a 4% (w/v) aqueous solution of Procion

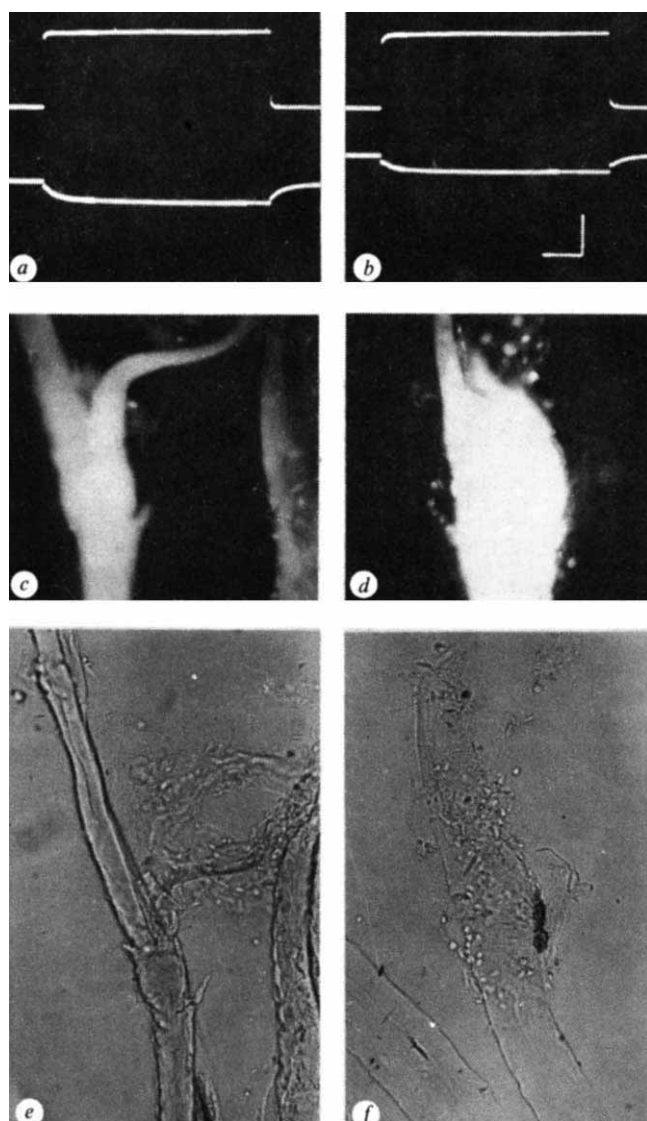
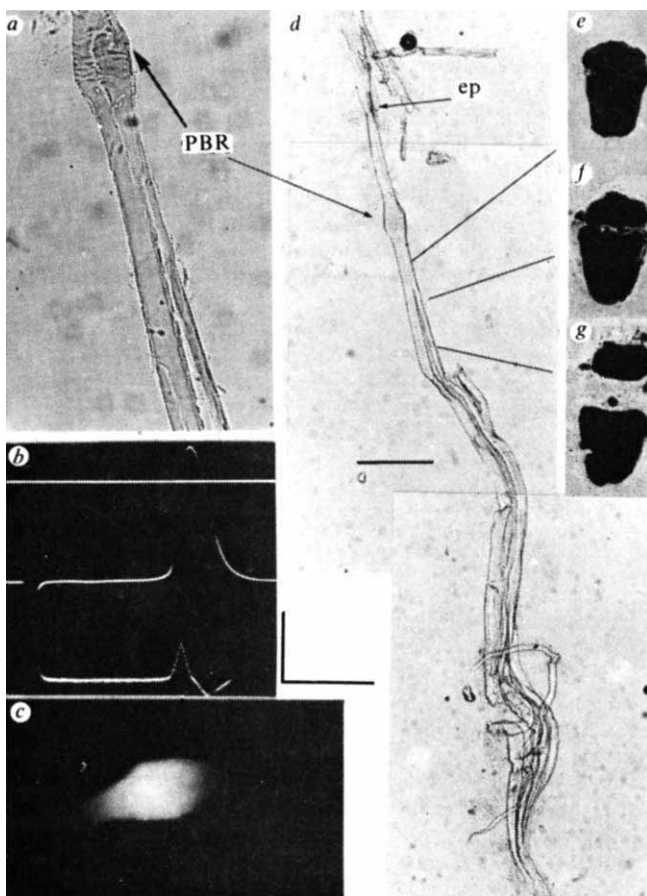


Fig. 1 Two typical experiments in which the input resistance of a muscle fibre was recorded using a standard two microelectrode technique. The current passing electrode was filled with a 4% solution of Procion brilliant red. The hyperpolarisation of the membrane (lower trace) in the two fibres following the passage of current (upper trace) is shown in *a* and *b*. The muscle fibres were then marked with dye by passing steady d.c. current for 10-20 s. The marked fibres were located using fluorescence microscopy (*c*, *d*) stained for cholinesterase and teased from the whole muscle (*e*, *f*). The first fibre (*a*, *c*, *e*) showed longitudinal splitting, and dye injection had taken place at the site of splitting. The second fibre (*b*, *d*, *f*) was 'short' and had probably undergone necrosis. M.e.p.ps were recorded from this fibre (not shown). Calibrations; *a*, *b*, 5 ms horizontal; 20 nA, 20 mV vertical; *e*, *f*, 70 μ m.

Fig. 2 An indirect action potential with its first derivative (*b*) was recorded from a dystrophic muscle fibre. The fibre was subsequently marked with Procion brilliant red and located using fluorescence microscopy (*c*). After staining for cholinesterase and teasing (*a*, *d*) the endplate cholinesterase (ep) and the site of the dye injection (PBR) were located. This fibre displayed longitudinal splitting, confirmed by serial section of the embedded fibre (*e*, *f*, *g*). Calibrations; *b*, 50 mV, 500 Vs⁻¹ vertical; 5 ms horizontal; *d*, 80 μ m.



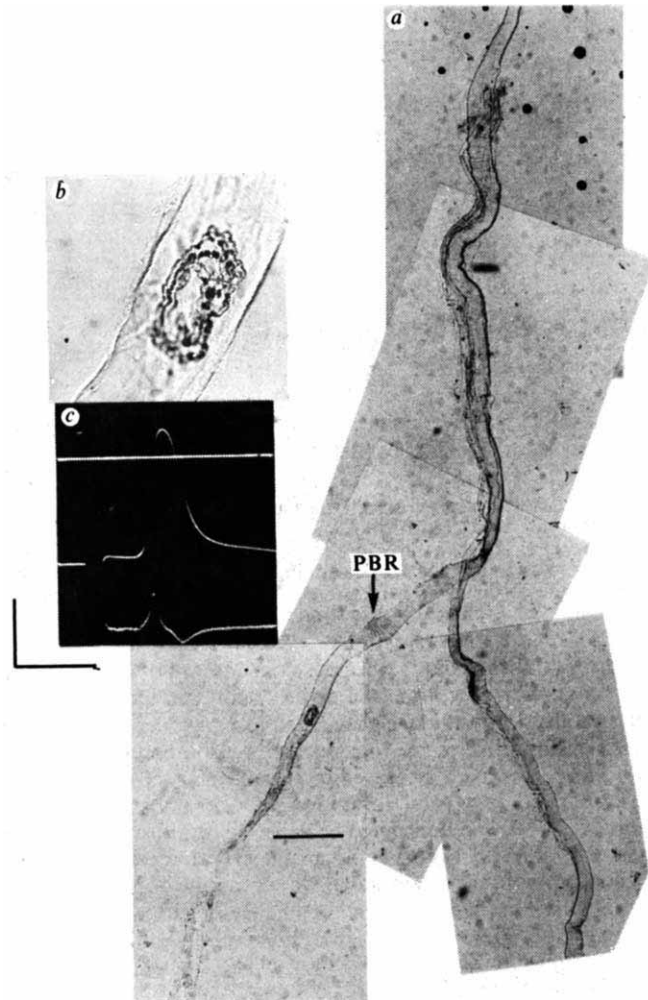


Fig. 3 An indirect action potential with its first derivative (*c*) was recorded from a dystrophic muscle fibre which was marked with dye (PBR in *a*). This fibre was very short, and it terminated before the normal tendon of insertion (*a*). Its cholinesterase distribution appeared normal (*b*). Calibration; *c*, 50 mV, 500 Vs⁻¹ vertical; 3 ms horizontal; *a*, 80 μ m.

brilliant red H3BN (ICI, Blackley) was introduced into the muscle fibre within 100 μ m of the recording electrode. Dye was ejected by passing inward direct current (20–80 nA; 10–20 s) through the dye-filled microelectrode. A visible red spot and a localised swelling of the muscle fibre indicated successful staining. Muscles were immediately fixed in cold (4 °C) formal-calcium (10% formalin, 1% CaCl₂) for 4 h, teased into small bundles and stained for cholinesterase activity⁸. The marked fibres were located using a Vickers Photoplan fluorescence microscope. The exciting light was filtered through a combination of Schott UG3, Schott BG12 and Balzers TRITC-I filters; emitted light was collected using a combination of Schott GG4 and Wratten 2E barrier filters. The marked fibres, which fluoresced bright red, were teased from the bundles, and, after photography, were embedded in 1% agar-agar, post-fixed in phosphate-buffered 4% glutaraldehyde (1 h) followed by phosphate-buffered 1% osmium tetroxide (15 min), and were then routinely processed and embedded in Spurr resin for subsequent microtomy. Sections were cut at 1 μ m and were stained with hot toluidine blue.

The experiments indicated that grossly abnormal fibres in the dystrophic muscle were sampled using the described techniques, and two of these fibres are shown in Fig. 1. In 12 fibres indirect action potentials were generated and the fibres were subsequently marked with dye and teased out of the muscle. Ten of the fibres showed abnormalities such as

Table 1 Mean quantum content of e.p.ps evoked at 3 Hz and 30 Hz nerve stimulation in uninjected fibres and fibres injected with Procion brilliant red

	Mean quantum content	
	3 Hz	30 Hz
Normal EDL (unmarked)	306.3 $\pm$ 22.5 (14)	184.2 $\pm$ 15.8 (14)
Normal EDL (marked)	346.2 $\pm$ 51.9 (5)	232.0 $\pm$ 35.4 (5)
Dystrophic EDL (unmarked)	381.9 $\pm$ 47.1 (22)	233.4 $\pm$ 24.4 (22)
Dystrophic EDL (marked)	379.0 $\pm$ 127.4 (12)	243.9 $\pm$ 40.9 (12)

Fibres were injected from microelectrodes filled with a 4% solution of dye. Figures are mean $\pm$ s.e.m. Mean values at corresponding frequencies are not significantly different ($P > 0.05$; Welch test).

longitudinal splitting^{9,10} (Fig. 2) or necrosis (Fig. 3). In a further 12 muscle fibres, the quantum content of the evoked e.p.ps was shown to be normal (Table 1); yet nine of the fibres show overt morphological abnormalities (Fig. 4 is typical).

We have thus demonstrated normal transmission in 24 dystrophic muscle fibres, 19 of which were grossly abnormal, exhibiting between them longitudinal fibre splitting, necrosis and, apparently, termination within the bulk of the muscle rather than at the normal myotendinous junction¹¹. We suggest, therefore, that the pathological involvement of muscle fibres in this disease cannot be either a direct or indirect consequence of a peripheral block of neuromuscular transmission.

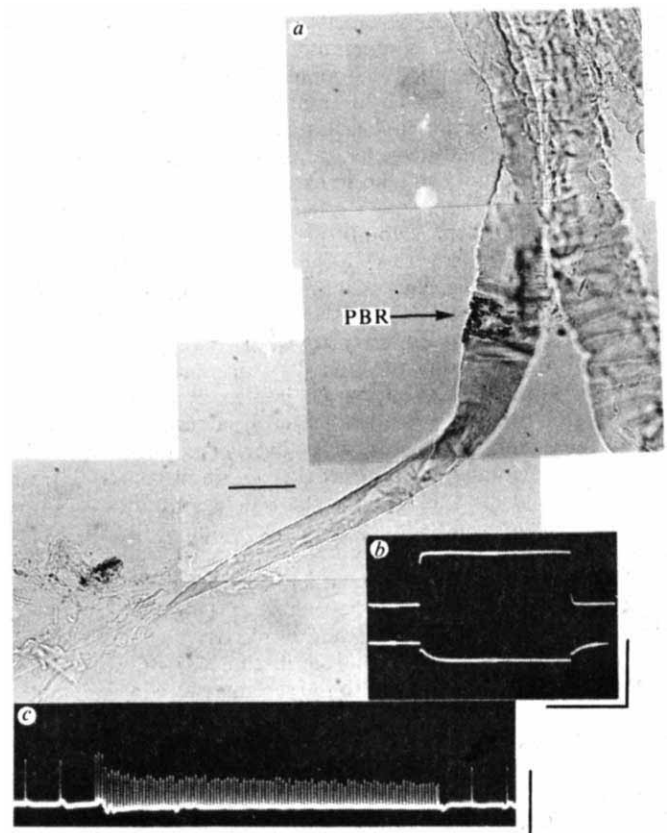


Fig. 4 Input resistance (*b*) and a train of endplate potentials (30 Hz, *c*) were recorded in the presence of *d*-tubocurarine from a dystrophic muscle fibre which was marked with dye (PBR in *a*). This fibre terminated well before the normal insertion of adjacent fibres. The cholinesterase staining was somewhat fragmented. The quantum content of the e.p.ps was normal (see Table 1). Calibration *a*, 50 μ m; *b*, 60 nA, 60 mV vertical, 15 ms horizontal; *c*, 10 mV vertical, 1 s horizontal.

We acknowledge the support of the Muscular Dystrophy Group of Great Britain and the MRC. Richard Ribchester holds an MRC Studentship.

J. B. HARRIS
R. R. RIBCHESTER

Muscular Dystrophy Research Laboratories,
Newcastle General Hospital,
Westgate Road,
Newcastle upon Tyne, UK

Received 12 August; accepted 20 November 1977.

- McComas, A. J. & Mrozek, K. *J. Neurol. Neurosurg. Psychiat.* 30, 526-530 (1967).
- Law, P. K., Atwood, H. L. & McComas, A. J. *exp. Neurol.* 51, 434-443 (1976).
- Harris, J. B. & Marshall, M. W. *Expl Neurol.* 41, 331-344 (1973).
- Harris, J. B. & Ribchester, R. R. *J. Physiol., Lond.* 263, 118-119P (1976).
- Kater, S. B. & Nicholson, C. (eds) *Intracellular Staining in Neurophysiology* (Karger, Basel, 1973).
- Harris, J. B. & Ribchester, R. R. *Br. J. Pharmac.* 55, 306-307P (1975).
- Martin, A. R. *J. Physiol., Lond.* 130, 114-122 (1955).
- Karnovsky, M. J. & Roots, L. *J. Histochem. Cytochem.* 12, 219-221 (1964).
- Michelson, A. M., Russell, E. S. & Harman, P. J. *Proc. natn. Acad. Sci. U.S.A.* 41, 1079-1084 (1955).
- Isaacs, E. P., Bradley, W. G. & Henderson, G. J. *Neurol. Neurosurg. Psychiat.* 36, 813-819 (1973).
- Williams, P. E. & Goldspink, G. J. *Anat.* 122, 455-465 (1976).

Neuromuscular blockade increases motoneurone survival during normal cell death in the chick embryo

NATURALLY occurring neuronal cell death is a quantitatively significant event in many developing systems¹⁻⁵; however, the determinants responsible for some neurones surviving while an equal or even greater number degenerate are still unknown. Neurones which undergo spontaneous degeneration have axons in the appropriate peripheral area^{6,7}, and it is generally believed that those neurones which die do so because they are losers in a competition for a limited number of 'targets'^{1,8}, or a limited amount of 'trophic substance'^{9,10}. This study was undertaken in an attempt to better understand the parameters which define a 'target' and the factors involved in the survival and degeneration of motoneurones in the lateral motor column of the lumbar spinal cord of the chick. Normally by embryonic day 6 a peak of approximately 22,000-24,000 motoneurones is found in each lateral motor column whereas by day 9 there are 13,500 and by day 12, and continuing through hatching, each lateral motor column consists of 11,000-11,500 motoneurones (unpublished data, also see refs 3, 11). We report here that contrary to what might be expected¹², embryos treated during the period of normal cell death with botulinum toxin (BTX) or curare contain about 5,500 more motoneurones (a 50% increase) in each lateral motor column after the period of cell death is over. Treatment with the irreversible nicotinic receptor blocker α -cobratoxin (CT, from *Naja naja siamensis*) also produced an increase in motoneurone survival; however, no studies were carried out with this agent beyond day 10.

Drugs were administered either by intramuscular (i.m.) injection (0.05-0.25 μ l) into the right leg (ventral muscle mass) on day 6 using a 33 gauge Hamilton microsyringe fitted with a 40- μ m glass tip¹³ (CT, BTX, and sterilised BTX), or by dropping the solutions (0.1-0.25 ml) on to the vascularised chorioallantoic membrane (curare and CT) on days 6, 7, 8, and 9. Embryos were killed for cell counts either on day 10 or day 16. After treatment with pre- or postsynaptic blocking agents a single leg at day 16 was capable of maintaining 50% more motoneurones (Fig. 1, Table 1). This increase did not seem to be a transient effect, as there were comparable numbers of neurones present in treated embryos on days 10 and 16. Moreover, no increase in neurone survival was found in embryos that were injected with curare after the period of cell death was over (group curare II in Table 1). Also, there was no enhancement of neuronal survival when drug treatment was not effective for the entire period between days 6 and 10 (group CT II).

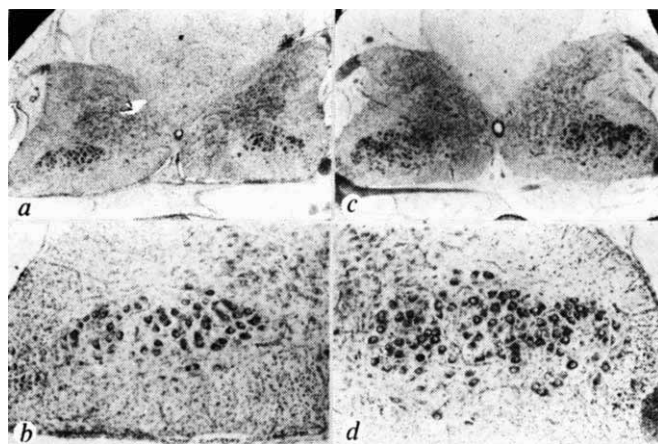


Fig. 1 Cross section through the spinal cord (fourth lumbar segment) of 16-d control (a, 20 $\times$; b, 51 $\times$) and curare-treated (curare I; c, 20 $\times$; d, 51 $\times$) embryos. The lateral motoneurones are the cluster of large dark staining cells in the ventro-lateral part of the cord. Spinal cords from BTX-treated embryos appear very similar to those from curarised embryos.

Curare, CT and BTX all decrease the amount of acetylcholine interacting with receptors, thereby reducing muscle contraction and motility. Motility levels are, therefore, a rough index of the drug's ability to interfere with normal neuromuscular function. All treatments which increased motoneurone survival produced either a partial or total inhibition of neuromuscular activity from day 6-10, and continued to decrease motility at least until day 16 (Table 1). Embryos in group CT II received a single injection of cobratoxin on day 6 which totally stopped all overt motility for the first half of the cell death period (day 6-7 $\frac{1}{2}$), whereas after day 8 motility levels were similar to controls. Interestingly, these embryos showed no increase in motoneurone numbers after killing on day 10. This suggests that either the critical period for increased cell survival occurs after day 8 or that once activity returns to near normal levels on day 8-8 $\frac{1}{2}$, 'redundant' synapses are lost, leading to a degeneration of the excess neurones.

The additional motoneurones present on day 16 seem to be uniformly distributed throughout the lateral motor column except for the most rostral segments (Fig. 2)—this distribution is very similar to the spatial distribution of 'hypothanasia' seen after transplanting additional limbs on to the body wall¹⁴. This may imply that virtually all muscles in the leg are capable of maintaining additional motoneurones. Other interpretations

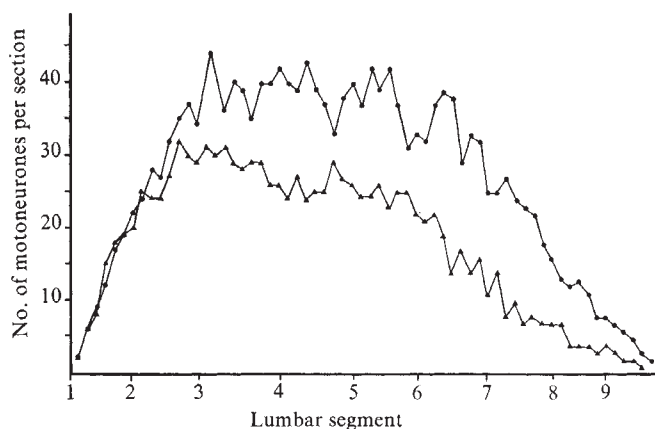


Fig. 2 Motoneurone counts in the right lateral motor column in lumbar segments 1-9 of 16-d control (▲) and curare-treated (●, curare I) embryos. Counts for every tenth section are represented. Note the increased cell number in segments 3-9 for curare-treated embryos. Motoneurone distribution after BTX treatment is basically the same as that after curare.

Table 1 Number of motoneurons and levels of embryonic activity after treatment with curare, botulinum toxin (BTX) and α -cobratoxin (CT)

Group* (age when killed)	Treatment	Motoneurons		% Change in motoneurone no.	Motility (moves min ⁻¹)				
		right LMC	left LMC		7	8	Day 9	10	16
10 d control (5)		13,720 $\pm$ 585	13,656 $\pm$ 670	—	6.8 $\pm$ 1.9	10.2 $\pm$ 1.7	18.1 $\pm$ 1.3	16.4 $\pm$ 1.9	—
10 d sterilised BTX(6)	i.m., day 6	13,816 $\pm$ 531	13,678 $\pm$ 688	—	—	—	—	13.3 $\pm$ 6.3	—
10 d curare(3)	2.5 mg on to CAM days 6, 7, 8, 9	17,067 $\pm$ 127	—	+24	4.1 $\pm$ 0.5	2.6 $\pm$ 0.5	0.4 $\pm$ 0.8	7.0 $\pm$ 3.0	—
10 d BTX(4)	i.m. day 6; 2,000 LD ₅₀	—	17,080 $\pm$ 704	+25	—	—	—	2.3 $\pm$ 0.7	—
10 d CT I(3)	100 μ g, day 6; 85 μ g days 7, 8; 50 μ g day 9, on to CAM	16,657 $\pm$ 710	—	+21	0	1.0 $\pm$ 0.7	0	5.0 $\pm$ 3.0	—
10 d CT II(8)	i.m. day 6; 6 μ g	13,398 $\pm$ 836	13,387 $\pm$ 816	-2 (both sides)	—	—	—	15.0 $\pm$ 2.1	—
16 d control(4)	—	11,360 $\pm$ 764	11,203 $\pm$ 577	—	—	—	—	—	14.6 $\pm$ 6.0
16 d BTX(5)	i.m. day 6; 1,200 LD ₅₀	—	16,402 $\pm$ 438	+46	—	—	—	—	0
16 d Curare I(3)	2.5 mg on to CAM day 6, 7, 8, 9	17,597 $\pm$ 1400	—	+55	1.8 $\pm$ 1.3†	4.1 $\pm$ 2.6	2.6 $\pm$ 0.7	3.6 $\pm$ 1.1‡	7.2 $\pm$ 2.5
16 d Curare II(3)	4.0 mg on to CAM days 12, 13 14, 15	11,537 $\pm$ 923	—	+2	—	—	—	—	0

Data are expressed as the mean $\pm$ s.d. Abbreviations: LMC, lateral motor column; CAM, chorioallantoic membrane; LD₅₀, 50% lethal dose for a 20g mouse. Weights of shank muscles (g) at day 16 were: control, 0.38 $\pm$ 0.04; BTX, 0.27 $\pm$ 0.02; curare, 0.26 $\pm$ 0.05. Lumbar cords were run up for paraffin histology, serially sectioned at 12 μ m, stained in thionine and the motoneurons in either one or both LMCs were counted in every tenth section^{3,11}. For groups in which counts of only one LMC are presented, at least one animal in each group was counted on both sides to assure that no discrepancies existed between right and left LMCs following treatment. The injection procedure itself did not produce any noticeable motoneurone degeneration (see groups CT II, and sterilised BTX); however, as an added precaution, only the left LMC was counted in BTX embryos. Motility recordings for CT I and curare-treated embryos between day 6 and 10 were made just before the daily drug treatment (that is, 24 h after the last treatment). Previous observations¹⁹ have shown that 2.5 mg of curare stops all activity for much of the 24-h period, followed by a subsequent slow rise in motility. Therefore, the activity levels presented represent the maximum activity for the preceding 24-h period.

*Number in group given in parentheses.

†Sample size for motility recordings on days 7–15 was $n = 5$; for day 16, $n = 8$.

‡Embryos were also recorded between days 10 and 16 but for illustrative clarity the data are not shown here. The average rate of motility during this time was about 8 per min compared to 16–22 per min for controls.

are possible, however, such as a preferential alteration of either fast or slow muscle properties and/or an increased number of axons forming synapses outside their normal area of innervation.

Our data show that decreasing functional activity at the neuromuscular synapse is correlated with an increase in the survival of motoneurons normally destined to undergo natural cell death. As has been previously reported^{15,16} curare, BTX and CT produced muscular atrophy; therefore, the possibility that an increase in the size or number of muscle fibres is responsible for the maintenance of the additional neurones is unlikely. It seems most likely that muscle fibres, which in normal conditions would maintain only one neurone, are capable of accepting and maintaining additional neurones after treatment. One explanation for this may be that normally the first axon to contact the myotube alters the membrane-receptor characteristics such that those parts of the myotube membrane not adjacent to the nerve terminal are rendered refractory to other axons. By blocking this initial nerve-muscle interaction we may have altered this process so that the myotube becomes receptive to additional axons^{17–19}. Alternatively, it might be that once the axon contacts the myotube and initiates contractions, there is a period in normal development during which the muscle membrane properties (ACh receptor number or distribution?) are changing as a result of the increased activity. Before and during this critical period, more than one neurone could be maintained by each myotube; however, once the number of contractions reached a critical level, the membrane properties would be altered such that only one neurone could be supported, and additional contacts would be lost. By decreasing myotube contractions with curare, BTX or CT, the critical period would be maintained with a resulting increase in motoneurone survival. The normal number of motoneurons present on day 10 in embryos which had a motility deficit before but not after day 8 (group CT II) could be accounted for by either model. Preliminary studies

with embryos in which motility is depressed from day 6–10 by daily injections of cobratoxin indicate that once motility returns to control levels (day 12–13) the additional neurones that are present on day 10 are lost; a finding which is also consistent with both models. The critical difference between the two models is that in the latter, but not the former, contact of the myotube by more than one axon is postulated to be a normal, though transient, process. We are carrying out experiments that may distinguish between these two alternatives.

Although our results support the idea that it is a limiting of 'targets' rather than of 'trophic substance' that is the important factor in naturally occurring cell death, in the future it may be neither profitable nor possible to draw any useful distinction between these two. We therefore emphasise that the novelty of our results is that they strongly implicate physiological activity at the neuromuscular synapse in the control of cell survival. Our results also support the suggestion that neurones which undergo normal cell death may form connections before spontaneous degeneration^{1,3}.

This work was supported by the North Carolina Division of Mental Health Services, Research Section and NSF grant BMS 75-23327. We thank Carol Willard and I.-W. Chu-Wang for their assistance. Botulinum toxin was a gift from Dr E. J. Shantz.

R. H. PITTMAN

R. W. OPPENHEIM

Neuroembryology Laboratory,
N.C. Division of Mental Health Services,
Research Section, Box 7532,
Raleigh, North Carolina 27611

Received 6 September; accepted 21 November 1977.

1. Cowan, W. M. in *Development and Aging in the Nervous System* (ed. Rockstein, M.) 19–41 (Academic, New York, 1973).
2. Landmesser, L. & Pilar, G. *J. Physiol., Lond.* **241**, 737–749 (1974).
3. Hamburger, V. *J. comp. Neurol.* **160**, 535–546 (1975).
4. Rubel, E. W., Smith, D. J. & Miller, L. C. *J. comp. Neurol.* **166**, 469–490 (1976).
5. Sohal, G. S. & Holt, R. K. *Expl Neurol.* **54**, 533–545 (1977).
6. Clarke, P. G. H. & Cowan, W. M. *J. comp. Neurol.* **167**, 143–164 (1976).
7. Oppenheim, R. W. & Chu-Wang, I. W. *Brain Res.* **125**, 154–160 (1977).

8. Landmesser, L. & Pilar, G. J. *Cell Biol.* **68**, 357–374 (1976).
9. Hamburger, V. *Am. J. Anat.* **102**, 365–410 (1958).
10. Prestige, M. C. in *The Neurosciences, Second Study Programme* (ed. Schmitt, F. O.) 73–82 (Rockefeller University Press, 1970).
11. Chu-Wang, I.-W. & Oppenheim, R. W. *J. comp. Neurol.* **177**, 33–58 (1978).
12. Changeux, J. P. & Danchin, A. *Nature* **264**, 705–712 (1976).
13. Oppenheim, R. W. & Heaton, M. B. *Brain Res.* **98**, 291–302 (1975).
14. Hollyday, M. & Hamburger, V. *J. comp. Neurol.* **170**, 311–320 (1976).
15. Drachman, D. B. *Archs Neurol., Chicago* **17**, 206–218 (1967).
16. Giacobini, F., Filogamo, G., Weber, M., Boquet, P. & Changeux, J. P. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1708–1712 (1973).
17. Gordon, T., Perry, R., Tuffery, A. R. & Vrbová, G. *Cell Tiss. Res.* **155**, 13–25 (1974).
18. Bennett, M. R. & Pettigrew, A. G. *J. Physiol., Lond.* **241**, 547–573 (1974).
19. Oppenheim, R., Pittman, R., Gray, M., & Maderdrut, J. *J. comp. Neurol.* (in the press).

Gating properties of acetylcholine receptor in newly formed neuromuscular synapses

A FOREIGN motor nerve transplanted on to an adult innervated muscle will not form functional neuromuscular synapses¹, yet several experimental procedures, such as denervation or muscle crush, change the properties of the muscle membrane so that a foreign nerve can establish 'ectopic' synapses in a normally endplate-free region of the muscle^{2–4}. The nerve thereby induces a localised area of high acetylcholine receptor (AChR) density in the extrajunctional membrane⁵. Denervated or previously crushed muscle fibres already have, before synapse formation, low density AChRs in the extrajunctional membrane^{6,7}. The extrajunctional AChRs are different in a number of respects from those in the junctional membrane^{8–11}. The question is, therefore, whether the AChRs of an ectopic synapse have properties similar to junctional or to extrajunctional AChRs.

The two types of AChR differ in their gating properties, since the mean open time of extrajunctional channels is about 3 to 5 times that of junctional channels^{12–14}. We have therefore measured the mean open time and the mean conductance of ion channels of ectopic synapses in frog (*R. temporaria*) sartorius muscle by analysing acetylcholine (ACh)-induced membrane current noise^{11,15}. The results indicate that the gating properties of ion channels in ectopic synapses are similar to those of normal synapses.

The tibialis nerve was connected to the tibial endplate-free zone of sartorius muscle. At the same time the muscle was cut or crushed near the site of implantation^{3,4}. In the majority of experiments 2 to 3 months later a connective septum has formed at the site of muscle crush, isolating muscle fibres of about 3–5-mm length extending between the septum and the tibial myotendinous junction. Most of these fibres respond to electrical stimulation of the tibialis nerve with an action potential or subthreshold endplate potentials (e.p.ps), and in some fibres low frequency (< 1 min⁻¹) miniature endplate potentials (m.e.p.ps) can be recorded^{3,4}. Occasionally, doubly innervated fibres are encountered which can be stimulated both by the tibialis and the sartorius nerve.

Examination of ACh sensitivity of tibialis-innervated fibres by focal iontophoretic application of ACh along a fibre reveals a peak of sensitivity at the site of the ectopic synapse. When these muscles are exposed to ¹²⁵I- α -bungarotoxin (BuTX), in most fibres several patches of high grain density are seen in the autoradiogram. The shape of these patches is round or elongated and they range in size from 5 to 30 μ m. In some fibres where elongated patches are present, few or no grains at all are seen on the rest of the fibre (Fig. 1a). In other fibres the rest of the membrane is also strongly labelled. Similar patches of high grain density are also found in muscles in which the tibial nerve had failed to establish synapses, for example, as a result of nerve displacement after the operation (Fig. 1b). They probably correspond to 'hot spots'^{16,17} found also in chronically denervated frog cutaneous pectoris muscle (A. Michler, personal communication). At present, we have no clue as to which of the patches found in innervated muscles corresponds to an ectopic endplate and which to a 'hot spot'.

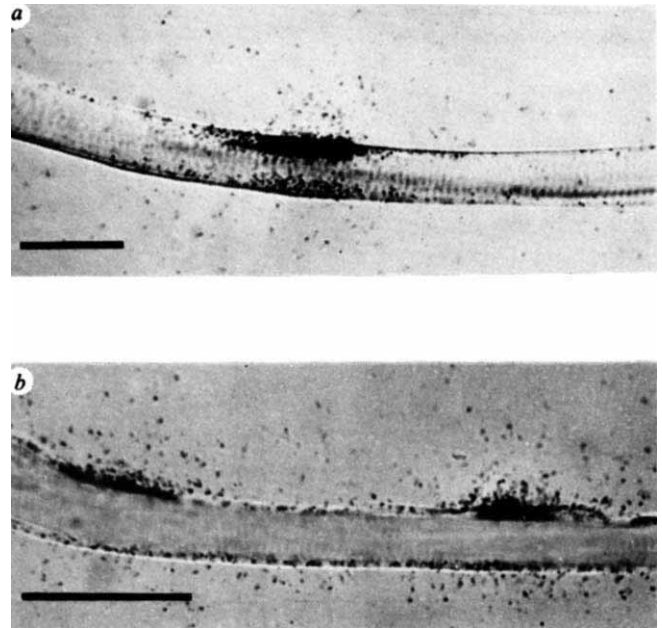


Fig. 1 Autoradiograms of muscle fibres isolated from the tibial portion of operated sartorius muscle. At the time of initial operation, the muscle was cut 5–10 mm away from its tibial insertion and the tibialis nerve was implanted into the remaining tibial portion. Several months later, muscles were exposed for 2 h to 1 μ g ml⁻¹ ¹²⁵I- α BuTX (specific activity: 4.7×10^5 Ci mol⁻¹) and subsequently fixed in glutaraldehyde. Single fibres were teased and dipped in Ilford L4 emulsion and exposed for 24 h. *a*, Elongated patch of high grain density in a fibre from a muscle which responded to stimulation of the tibialis nerve. Experiment 116 d after operation. *b*, Patches of high grain density ('hot spots') in a fibre from a muscle in which the tibialis nerve did not make contact as a result of mechanical displacement. The fibres were in effect chronically denervated for 89 d. Scale bar, 20 μ m.

A tentative hypothesis is that the elongated patches in fibres with otherwise low grain density (Fig. 1a) represent ectopic endplates.

To study their channel properties, superficial ectopic endplates were voltage clamped to a holding potential between -70 mV and -90 mV. Membrane currents of 15 to 130 nA mean amplitude were induced by releasing ACh from a micropipette located 30–50 μ m above the synapse. During ACh application, the clamp current fluctuates around its mean in a manner similar to that described for normal endplates. The autocorrelation function of ACh-induced current noise¹² shows a single decay constant τ_{noise} , the value of which varied between 1.36 ms and 1.85 ms in seven experiments carried out at 15–18 °C, the mean value being $\tau_{\text{noise}} = 1.5 \pm 0.3$ ms ($\pm$ s.e.m.). The values of individual experiments were normalised to -80 mV membrane potential and 18 °C assuming¹⁵ a Q_{10} of 2.77 and an e -fold increase of τ per 80 mV hyperpolarisation. The mean value of τ_{noise} which is an estimate of the channel open time^{11,15} is thus very similar to that reported for normal synapses in *R. temporaria*¹². The result of one experiment is shown in Fig. 2a.

The size of the single channel current i obtained from the ratio of the variance σ^2 of the current fluctuations to the mean current I was more variable and ranged between 0.7 pA and 1.7 pA in these experiments. The reversal potential of nerve-evoked endplate currents (e.p.cs) in ectopic synapses was determined in three experiments and was found to be between -5 mV and 0 mV (Fig. 3). This is very similar to what has been reported as a reversal potential for e.p.cs evoked by nerve stimulation or iontophoretic ACh application at normal synapses^{18,19}. The size of the elementary conductance change has a mean value of $\gamma = 16 \pm 1.4$ pS (range 10–21 pS, seven experiments, reversal potential assumed at 0 mV) and is thus somewhat lower than the value obtained previously for normal synapses¹².

A possible reason for the lower estimate of γ in ectopic synapses as compared to normal synapses could have been that during ACh application an appreciable fraction of all available receptors is liganded, since AChE activity is low in these synapses. Therefore, the low concentration limit required for estimating i from the ratio σ^2/I might not hold and i would be underestimated²⁰. This was tested in one experiment by inducing mean membrane currents which varied in size over a fourfold range. In this experiment, the estimate of γ was independent of the magnitude of the mean current. Another possibility for the smaller estimate of γ might be that in most experiments a relatively high concentration of Mg^{2+} was present in the bathing solution.

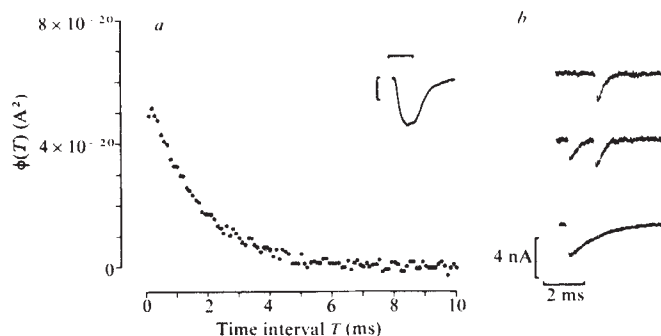


Fig. 2 Analysis of endplate currents at ectopic synapse. All experiments were carried out in frog Ringer solution of the following composition (in mM): NaCl, 114; KCl, 2.5; CaCl₂, 0.5–1; MgCl₂, 0–15; Na₂HPO₄, 2.15; NaH₂PO₄, 0.5; pH, 7.1. MgCl₂ was added in five out of seven experiments to abolish nerve-evoked muscle contraction. Superficial ectopic synapses were located by the rise time of e.p.ps 97 and by applying hypertonic sucrose from a blunt micropipette along the fibre. A brisk increase in m.e.p.p. frequency indicated the site of innervation. *a*, The graph shows the autocorrelation function of ACh-induced voltage clamp current fluctuations. Data points represent value of autocorrelation function $\Phi(T)$ for different time intervals T . Time course of endplate current during iontophoretic ACh application is shown in the inset (calibration marks indicate 20 nA and 25 s). A current sample of 6.4 s duration, taken during the peak of the response, was used for calculation of autocorrelation function. Mean amplitude: $I = 43$ nA. The record was filtered with a band pass (0.5 Hz–2.5 kHz). Sampling interval 100 μ s. Control data obtained from a current sample in the absence of ACh was subtracted. Continuous line is an exponential function fitted by eye to the data points. It yields a value of $\tau = 1.6$ ms for the time constant and an extrapolated ordinal intercept of $\sigma^2 = 6.1 \times 10^{-20} A^2$ for the variance of ACh-induced current fluctuations. Holding potential: -80 mV. Temperature $18^\circ C$. Experiment performed 104 d after implantation of the tibialis nerve. *b*, Upper two traces: Examples of m.e.p.s. recorded from the same fibre as in (*a*), $\tau_{m.e.p.c.} = 3.4$ ms. Bottom trace: m.e.p.c. from another fibre with extremely long decay time $\tau_{m.e.p.c.} = 12$ ms at -80 mV membrane potential and $18^\circ C$. Time constant of current fluctuations in this fibre was $\tau_{noise} = 1.8$ ms.

Whereas the elementary conductance events in normal and ectopic synapses are similar, the time course of miniature endplate currents (m.e.p.c.s) is different in the two types of synapses. In a normal synapse the m.e.p.c. decay time constant $\tau_{m.e.p.c.}$ is similar or identical to the mean open time of ion channels as derived from noise analysis¹⁵. In all ectopic synapses studied, $\tau_{m.e.p.c.}$ was longer by a factor of 2 to 6 than τ_{noise} and ranged from 3.3 ms to 12.6 ms (see Fig. 2*b*, Table 1). In contrast to the relative uniformity of τ_{noise} in different fibres, the value of $\tau_{m.e.p.c.}$ changed from fibre to fibre. This is similar to what is observed at normal endplates after inhibition of AChE, suggesting that the AChE activity in ectopic synapses is low compared to that in a normal synapse. In most ectopic synapses we failed to detect AChE activity histochemically by the Karnovsky method²¹, even after several hours of incubation and despite intense staining of myotendinous junctions which were only a few millimetres away. We also found in three experiments that adding 3–6 μ M neostigmine

Fig. 3 Determination of reversal potential of nerve-evoked e.p.c. in a newly formed synapse. Membrane potential was clamped to the value indicated beside each trace and 20–50 responses were averaged at each potential level. Arrow indicates time of tibialis nerve stimulation. E.p.c. reverses sign between -5 mV and 0 mV membrane potential. Experiment was carried out 89 d after operation.

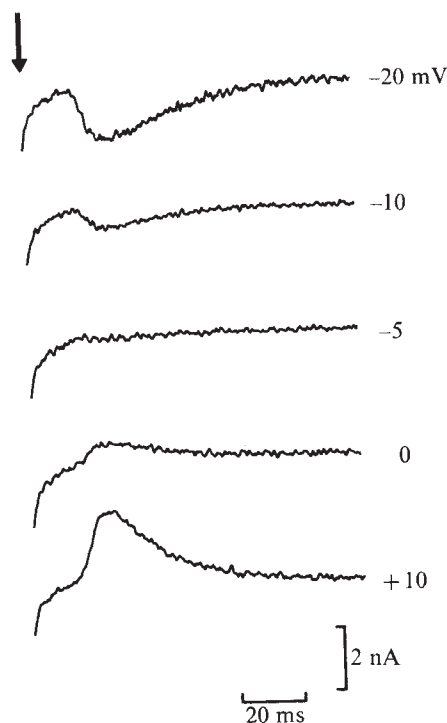


Table 1 Comparison of mean channel open time and mean m.e.p.c. decay time constants obtained from the same ectopic synapse

Experiment	τ_{noise} (ms)	$\tau_{m.e.p.c.} \pm$ s.e.m. (ms)	Membrane potential (mV)	Temperature ($^\circ C$)	Time after operation (d)
1	1.58	3.76 (0.09)	-80	17	90
2	1.82	5.14 (0.21)	-70	15	91
3	1.85	12.23 (0.35)	-80	18	93
*4	1.61	3.36 (0.29)	-80	18	115
*5	1.52	6.96 (0.35)	-80	18	113
6	1.71	—	-70	16	81
7	1.45	4.22 (0.15)	-90	16	97

$\tau_{m.e.p.c.}$ represents mean values obtained from 9 to 25 individual m.e.p.c.s. Time after operation refers to the time between implantation of the tibialis nerve and the time the experiment was carried out.

*Experiments done in frog Ringer solution with no Mg^{2+} ions present. In the remainder Mg^{2+} concentration varied between 5 and 15 mM.

to the bath prolonged $\tau_{m.e.p.c.}$ by a factor of less than 1.25 as compared to a factor of about 2 reported for normal neuromuscular synapses^{22,23}.

We have shown that ion channels associated with AChRs of newly formed ectopic synapses on adult muscle are similar in their gating properties to those found in normal endplates. They are different from those of extrajunctional AChRs found in denervated fibres not contacted by a nerve. The question arises as to whether, during synapse formation in adult fibres, extrajunctional AChRs are accumulated under the influence of the motor nerve, as has been shown recently for embryonic amphibian muscle²⁴ and are changed in their properties to junctional AChRs, or whether the motor nerve induces synthesis of 'junctional' receptors at its site of contact. Experiments carried out at early stages of synapse formation may help to answer this question.

We thank Professor J. Schmidt for ¹²⁵I- α -BuTX and Ms A. Michler for help with autoradiography. H. R. Brenner was supported by the Schweizerische Nationalfonds.

H. R. BRENNER
B. SAKMANN

Abteilung Neurobiologie,
Max-Planck-Institut für biophysikalische Chemie,
3400 Göttingen-Nikolausberg,
Am Fassberg, FRG

Received 28 October; accepted 1 November 1977.

1. Elsborg, C. A. *Science* **45**, 318–320 (1917).
2. Fex, S. & Thesleff, S. *Life Sci.* **6**, 635–638 (1967).
3. Miledi, R. *Nature* **193**, 281–282 (1962).
4. Miledi, R. *Nature* **199**, 1191–1192 (1963).
5. Lomo, T. & Slater, C. R. *J. Physiol., Lond.* **258**, 107 P. (1976).
6. Miledi, R. *J. Physiol., Lond.* **151**, 24–30 (1960).
7. Katz, B. & Miledi, R. *J. Physiol., Lond.* **170**, 389–396 (1964).
8. Berg, D. K. & Hall, Z. W. *Science* **184**, 473–475 (1974).
9. Chang, C. C. & Huang, M. C. *Nature* **253**, 643–644 (1975).
10. Beranek, R. & Vyskočil, F. *J. Physiol., Lond.* **188**, 53–66 (1976).
11. Katz, B. & Miledi, R. *J. Physiol., Lond.* **224**, 665–699 (1972).
12. Neher, E. & Sakmann, B. *J. Physiol., Lond.* **258**, 705–730 (1976).
13. Dreyer, F., Walther, Chr. & Peper, K. *Pflügers Arch. ges. Physiol.* **366**, 1–9 (1976).
14. Sakmann, B. *Pflügers Arch. ges. Physiol.* **359**, R89 (1975).
15. Anderson, C. R. & Stevens, C. F. *J. Physiol., Lond.* **235**, 655–691 (1973).
16. Sytkowski, A. J., Vogel, Z. & Nirenberg, M. W. *Proc. natn. Acad. Sci. U.S.A.* **70**, 270–274 (1973).
17. Ko, P. K., Anderson, M. J. & Cohen, M. W. *Science* **196**, 540–542 (1977).
18. Colquhoun, D., Dionne, V. E., Steinbach, J. H. & Stevens, C. F. *Nature* **253**, 204–206 (1975).
19. Mallart, A., Dreyer, F. & Peper, K. *Pflügers Arch. ges. Physiol.* **362**, 43–47 (1976).
20. Stevens, C. F. *Biophys. J.* **12**, 1028–1047 (1972).
21. Karnovsky, M. J. *J. Cell Biol.* **23**, 217–222 (1964).
22. Gage, P. W. & McBurney, R. N. *J. Physiol., Lond.* **244**, 385–407 (1975).
23. Hartzell, H. C., Kuffler, S. W. & Yoshikami, D. *J. Physiol., Lond.* **251**, 427–463 (1975).
24. Anderson, M. J. & Cohen, M. W. *J. Physiol., Lond.* **268**, 751–773 (1977).

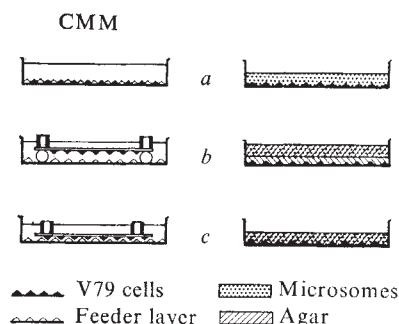


Fig. 1 Schematic representation of the experiments using CMM or MMM systems, which were designed either to allow or to prevent contact between the V79 cells and metabolic activation systems. Experimental procedures for CMM were adopted from Huberman and Sachs¹. Following irradiation at 4,000 R, secondary cultures of BDVI rat embryo cells were plated at 2.5×10^6 cells per 60-mm dish (feeder layer) and, 5 h later, 3.0×10^6 V79 cells were plated and cultured overnight in Eagle minimum essential medium plus 10% foetal calf serum (FCS). The next day, cultures were treated with 7,8-diol BP (obtained from Dr P. L. Grover, Chester Beatty Research Institute), dissolved in DMSO, at concentrations of 1 and 2 nmol ml⁻¹ (8 ml per dish) for 48 h, followed by plating for determination of induced cytotoxicity and mutagenicity. Experimental procedures for MMM have been described elsewhere²; in brief, the V79 cells (overnight culture after plating 1.5×10^6 cells per 60-mm dish) were incubated for 2 h in 2.5 ml of the reaction mixture, either in liquid suspension or suspended in 0.5% agar. The reaction mixture consisted of 0.75 ml post-mitochondrial fraction from phenobarbitone-pretreated BDVI rat livers (1 mg ml⁻¹ in drinking water for 7 d), 0.75 ml modified Sørensen phosphate buffer (0.055 M, pH 7.4, containing 0.9% NaCl and 1.6 mg ml⁻¹ MgCl₂·6H₂O) and 1.0 ml PBS containing 12.5 μ mol glucose-6-phosphate and 2 μ mol NADP⁺. The cells were then washed twice and incubated for 2–3 h in fresh culture medium, followed by plating for determination of induced cytotoxicity and mutagenicity. Cytotoxicity was determined by plating 100 cells per 60-mm dish (four dishes at each point) and culturing for 7 d. For mutagenesis, 2×10^4 and 10^5 cells per 60-mm dish were plated for *aza* and *oua* respectively (eight dishes at each point). After an expression period of 2 d, selection drugs were added to give final concentrations of 20 μ g ml⁻¹ 8-azaguanine or 1 mM ouabain. The media containing drugs were changed once, 5–7 d later; for the 8-azaguanine medium, FCS was replaced by dialysed FCS and non-essential amino acids were added at 0.1 mM. The cultures were fixed and stained with Giemsa 12 d (*aza*) or 14 d (*oua*) after plating. Mutation frequency was calculated per 10^6 survivors, taking into account the number of cells plated and the plating efficiency. The stability of *aza* and *oua* was demonstrated using 10 and 11 isolated colonies, respectively, which were cultured for more than one month in the absence of the drug used for selection³. *a*, Conventional CMM and MMM systems, as described above, in which the V79 cells are allowed to remain directly in contact with the cells or microsomes carrying metabolic activating enzymes. *b*, Is designed to separate the V79 cells from the metabolic activation system. In the CMM system, a glass cover slip (24 $\times$ 40 mm) on which 1.4×10^6 V79 cells were plated and cultured overnight, was supported above the feeder layer at a distance of approximately 1 mm. In the MMM system, 2 ml of 0.5% agar in PBS was poured on the V79 monolayer (approximately 1 mm in thickness) and, after hardening, the reaction mixture suspended in 0.5% agar was added. *c*, In the CMM system, the V79 monolayer was grown on a cover slip and placed directly on the feeder layer. In the MMM system, the reaction mixture was suspended in 0.5% agar and poured directly on the V79 monolayer.

Mutations in two genetic loci, comprising 8-azaguanine resistance (*aza*) and ouabain resistance (*oua*), were studied using V79 Chinese hamster cells. Mutations were induced either by *trans*-7,8-dihydroxy-7,8-dihydrobenzo[a]pyrene (7,8-diol BP) at concentrations of 1 and 2 nmol ml⁻¹ in the CMM system, or by *N*-nitrosodimethylamine (DMN) at concentrations of 10 and 30 μ mol ml⁻¹ in the MMM system. Both these chemicals require metabolic activation^{3,5}. As shown in Fig. 1, the following experiments were designed either to allow, or to prevent, contact between the V79 cells and the feeder layer, or the microsomes carrying the enzymes responsible for metabolic activation: *a*, conventional mutagenesis assays

Direct or proximate contact between cells and metabolic activation systems is required for mutagenesis

THE adverse biological effects of most carcinogens and/or mutagens are dependent on the formation of electrophilic intermediates which react with nucleophilic groups in cellular macromolecules. Inclusion of a metabolic activation system in various assays enables the detection of mutagenic activity for carcinogens and/or mutagens which require such metabolic activation. In mutagenesis assays, using mammalian cells, two metabolic activation systems have been used: cell-mediated mutagenesis (CMM), in which cells are co-cultured with metabolically competent, but lethally irradiated, cells (feeder layer)^{1,2}, and microsome-mediated mutagenesis (MMM), in which cells are treated with chemicals in the presence of a microsomal fraction from rodent livers and an NADH-generating system^{3–5}. The present study was carried out in order to investigate whether direct or proximate contact between target cells and such mediators of metabolism is necessary for the induction of mutagenesis. The results suggest that such contact is essential in both cell- and microsome-mediated mutagenesis of mammalian cells and, possibly, also in bacterial systems.

were used. In the CMM system the V79 cells were grown directly on the feeder layer of rat embryo cells and, in the MMM system, the cells were incubated in the presence of a reaction mixture containing a post-mitochondrial liver fraction from phenobarbitone-pretreated rats and co-factors; *b*, to separate the V79 cells from the feeder layer in the CMM system, a glass cover slip, on which a monolayer of the V79 cells had been grown, was placed above the feeder layer at a distance of approximately 1 mm. In order to maintain a constant distance between the two cell layers, the cover slip was supported by two strips of silicone rubber and pressed down with two metal cylinders used for colony isolation. In the MMM system, the V79 cells were separated by an agar layer (0.5% in phosphate-buffered saline, PBS) at a distance of approximately 1 mm from the reaction mixture, suspended in agar; *c*, this was designed as a control of the experimental conditions described in *b*. In the CMM system, the V79 monolayer, on a cover slip, was placed face down, directly on the feeder layer and was pressed down with two metal cylinders to ensure maximum contact between the two cell layers. In the MMM system, the reaction mixture suspended in agar was poured directly on to the V79 monolayer.

Figure 2 shows the mutagenicity of 7,8-diol BP in CMM; it produced *aza* and *oua* in a dose-related fashion when the V79 cells were cultured on the feeder layer in the conventional CMM assay (*a* in Fig. 1). No mutations were induced, however, either when the V79 cells were separated at a distance of approximately 1 mm from the feeder layer (*b* in Fig. 2), or when the V79 monolayer grown on the glass cover slip was placed directly on to the feeder layer (*c* in Fig. 1). In the latter, growth of the V79 cells was considerably reduced, presumably due to lack of circulation of nutrients under the cover slip; doubling time was more than 60 h compared with 15–16 h for the other systems (*a* and *b*). The results indicate that 7,8-diol BP, one of the metabolites of BP, is further metabolised in the CMM assay to an ultimate form, possibly 7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene (7,8-diol-9,10-oxide BP)⁶⁻⁹, which could be transferred to the target cells through direct or proximate contact with the feeder layer.

Similar results were obtained in the MMM assay (Fig. 3). Treatment of the cells with DMN in the presence of post-mitochondrial liver fraction and cofactors, either in liquid (*a* in Fig. 1), or in agar (*c* in Fig. 1), induced dose-related cytotoxicity and mutagenicity. However, neither cytotoxicity nor mutagenicity were induced when the cells were separated by the agar layer approximately 1 mm thick (*b* in Fig. 1). This lack of induction could be explained either by the requirement of direct contact or close proximity between the target V79 cells and microsomes carrying metabolising enzymes, or by the decomposition or trapping of methylating intermediates formed from DMN in the separating agar layer. The former possibility is further suggested by a conventional MMM assay in which the V79 monolayer was treated with DMN at 30 $\mu\text{mol ml}^{-1}$ in different volumes of the reaction mixture in

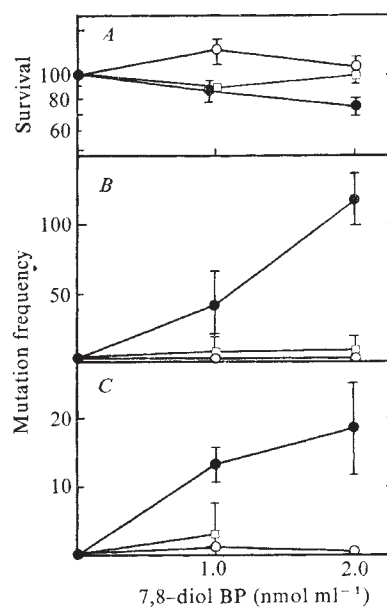


Fig. 2 Cytotoxicity (A) and mutagenicity (B, *aza*; C, *oua*) induced by 7,8-diol BP at concentrations of 1 and 2 nmol ml^{-1} in the CMM assay. ●, The conventional CMM assay (*a* in Fig. 1) in which the V79 cell grew directly on the feeder layer. ○, *b* In Fig. 1, in which the V79 cells were separated from the feeder layer. □, *c* In Fig. 1, in which the V79 monolayer was placed on the feeder layer. Cytotoxicity was expressed as percentages of plating efficiency of the solvent (DMSO) control. Mutation frequencies were expressed as the number of drug resistant clones per 10^5 survivors. *oua* In Fig. 1c at 2 nmol ml^{-1} was not examined. Bars indicate s.e.; when not indicated, s.e. are within the circles.

liquid (0.5, 1.5 and conventional 2.5 ml, which correspond approximately to 0.25, 0.75 and 1.25 mm in depth, respectively). The mutation frequencies obtained (mean number of resistant clones per 10^5 survivors, $\pm$ s.e.m.) were 173 ± 18 , 212 ± 26 and 171 ± 19 for *aza* and 41 ± 4 , 60 ± 5 and 50 ± 4 for *oua*, respectively indicating that induction of mutations may take place independently of the total volume, in the thin layer of the reaction mixture.

In the experiment shown in Table 1, V79 cells were treated with a directly acting mutagen, *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG), while separated from the MNNG-containing agar medium by an agar layer of varying thickness (approximately 0.5, 1.0 and 1.5 mm). No decreases in the induction of cytotoxicity and mutagenicity were observed when the cells were separated from the MNNG medium. There was also no difference in MNNG-induced mutation frequency between liquid and agar medium. These results indicate that agar did not trap or interfere with the diffusion of MNNG,

Table 1 MNNG-induced cytotoxicity and mutagenicity when the cells were separated from MNNG-containing agar medium by agar layers of varying thickness

Treatment		Agar layer for the separation		P.E. (%)	Mutation frequency	
		volume (ml)	thickness (mm)		<i>aza</i>	<i>oua</i>
DMSO	liquid	0	0	93.5±4.8	4.0±1.9	0
MNNG	liquid	0	0	66.0±2.9	46.6±3.6	4.7±1.0
MNNG	agar	0	0	68.3±7.7	46.6±5.8	4.6±1.2
MNNG	agar	1	0.5	58.3±4.4	47.3±6.5	5.2±1.3
MNNG	agar	2	1.0	65.8±4.0	50.4±5.4	4.8±1.4
MNNG	agar	3	1.5	51.8±2.0	66.7±9.1	11.1±1.6

V79 cells were plated at 1.5×10^6 cells per 60-mm dish, cultured overnight and then treated with MNNG at 4 nmol ml^{-1} (4 μl per dish), either in liquid medium, or in 0.5% agar medium, for 2 h. The cells were separated from the MNNG-containing medium by agar layers (0.5% in phosphate-buffered saline (PBS), approximately 0.5, 1.0 and 1.5 mm in thickness, corresponding to 1.0, 2.0 and 3.0 ml of agar suspension. The cultures were then washed three times and cultured in a fresh culture medium for 2–3 h, followed by plating for determination of induced cytotoxicity and mutagenicity (see the legend for Fig. 1). This replating procedure produced a lower mutation frequency, approximately 20% for *aza* and 8% for *oua*, as compared to the conventional *in situ* treatment.

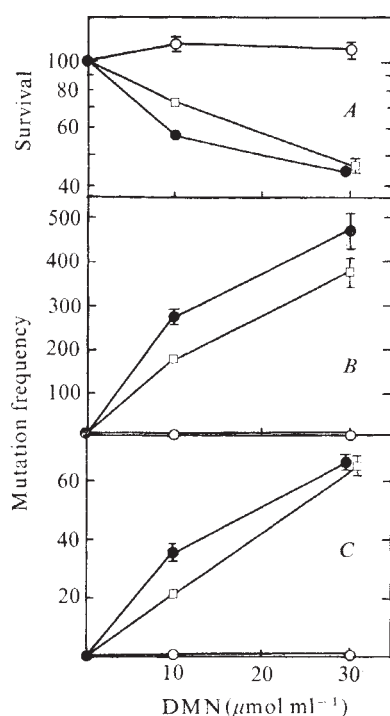


Fig. 3 Cytotoxicity (A) and mutagenicity (B, *aza*; C, *oua*) induced by DMN at concentrations of 10 and 30 $\mu\text{mol ml}^{-1}$ in the MMM assay. ●, The conventional MMM assay (a in Fig. 1) in which the V79 monolayer was incubated in the presence of the reaction mixture in liquid suspension. ○, b In Fig. 1, in which the V79 cells were separated by an agar layer from the reaction mixture suspended in agar. □, c In Fig. 1, in which the reaction mixture suspended in agar was poured directly on to the V79 monolayer. Cytotoxicity and mutagenicity are expressed as described in the legend to Fig. 2.

an alkylating agent that is relatively stable (half life is 90 min^{10,11}).

These results suggest that direct or proximate contact between the target cells and membranes of the cells or microsomal particles, carrying carcinogen-metabolising enzymes is essential in cell- and microsome-mediated mutagenesis of mammalian cells. This requirement may also hold true for mutagenesis assays using bacteria and microsomes. Glatt and Oesch¹², using electron microscopy, have demonstrated tight binding of microsomal particles to *Salmonella typhimurium* TA98, suggesting the importance of such a direct contact in bacterial systems.

Two mechanisms are plausible in explaining why direct or proximate contact is required. One is the chemical instability of the ultimate reactive metabolites formed, which, in many cases, are short-lived intermediates. For example, the half lives of synthetic 7,8-diol-9,10-oxide BP, an ultimate form of 7,8-diol BP, is 0.5 min for the *syn* isomer and 2.5 min for the *anti* isomer in aqueous media⁹; the half life of methylating intermediate(s) formed from DMN by liver microsomal mono-oxygenases was measured to be of the order of seconds (A. Barbin & H. Bartsch, personal communication). Such unstable compounds may thus decompose in aqueous media before reaching the DNA of the target cells, unless the distance of migration is reduced to a minimum. Second, when ultimate metabolites with lipophilic properties are formed, they could reach the target cells by transfer through the lipid matrix of membranes. This would require direct contact between the target cells and the metabolic activation systems. Such lipophilic environments may eventually contribute to a prolonged half-life of reactive compounds and may prevent contact with inactivating enzymes, such as glutathione S-transferases present in the cytosol¹².

We thank Drs H. Bartsch and R. Montesano and Messrs C. Malaveille and J. D. Wilbourn for their useful comments.

This work was supported in part by the US NCI (contract 1CP-55630).

T. KUROKI
C. DREVON

Unit of Chemical Carcinogenesis,
International Agency for Research on Cancer,
150 cours Albert Thomas,
69372 Lyon Cedex 2, France

Received 23 August; accepted 21 November 1977.

- Huberman, E. & Sachs, L. *Int. J. Cancer* **13**, 326–333 (1974).
- Newbold, R. F., Wigley, C. B., Thompson, M. H. & Brookes, P. *Mutat. Res.* **43**, 101–116 (1977).
- Umeda, M. & Saito, M. *Mutat. Res.* **30**, 249–254 (1975).
- Krahn, D. & Heidelberger, C. *Mutat. Res.* **46**, 27–44 (1977).
- Kuroki, T., Drevon, C. & Montesano, R. *Cancer Res.* **37**, 1044–1050 (1977).
- Huberman, E., Sachs, L., Yang, S. K. & Gelboin, H. V. *Proc. natn. Acad. Sci. U.S.A.* **73**, 607–611 (1976).
- Newbold, R. F. & Brookes, P. *Nature* **261**, 52–54 (1976).
- Wood, A. W. *et al. Cancer Res.* **36**, 3358–3366 (1976).
- Malaveille, C., Kuroki, T., Sims, P., Grover, P. L. & Bartsch, H. *Mutat. Res.* **44**, 313–326 (1977).
- Lawley, P. D. & Thatcher, C. J. *Biochem. J.* **116**, 693–707 (1970).
- Bertram, J. S. & Heidelberger, C. *Cancer Res.* **34**, 526–537 (1974).
- Glatt, H. R. & Oesch, F. *Archs Toxicol.* (in the press).

Ascorbate-enhanced cytotoxicity of misonidazole

MISONIDAZOLE, an electron-affinic, nitroheterocyclic compound (1-(2-nitro-1-imidazole)-3-methoxy-2-propanol, Ro-07-0582) is undergoing preliminary clinical trials, which are exploring its capacity to sensitise hypoxic tumour cells to ionising radiation damage^{1–4}. The drug also possesses a substantial cytotoxic effect, independent of radiation, which is selectively expressed in hypoxic cells^{5,6}. Both of these properties may contribute to the potential value of this agent as an adjunct to radiation therapy of malignant disease in those situations where radioresistant hypoxic cells represent the limiting factor in achieving local control. It is also possible that the cytotoxic property could be a useful adjunct to chemotherapy as it should be effective against noncycling hypoxic tumour cells. Misonidazole, however, may be cytotoxic to the normal hypoxic tissues in the human body (nerve, skin, cartilage, and so on) and this is a major concern in the clinical application of the drug^{1,4}. Furthermore, misonidazole is metabolised in mammalian cells and at least two products have been isolated⁷, toxic to both aerobic and hypoxic mammalian cells. Thus, if they are able to diffuse freely, the metabolites may also be a threat to aerobic tissues⁸. This cytotoxicity of misonidazole is temperature dependent^{5,9,10}, and varies with cell line^{11,12}, and cysteamine seems to protect against it¹³. It leads to strand breaks in cellular DNA and those cells which fail to survive also fail to repair these strand breaks¹². In this work we show the effect of ascorbic acid (vitamin C) on the cytotoxicity of misonidazole.

Chinese hamster ovary cells (CHO) were grown in suspension culture according to standard procedures⁵. Growth medium containing the drugs to be tested was de-oxygenated in a stirred glass vessel by flowing purified nitrogen over the solution for 10–15 min before the addition of hypoxic cells (2×10^5 cells ml^{-1}). The flow of nitrogen was maintained throughout the experiment. At the prescribed times, samples were withdrawn from the suspension, diluted, washed and plated in plastic culture dishes for the assay of colony forming ability⁵.

The effect of ascorbate on the hypoxic cytotoxicity of misonidazole is shown in Fig. 1. The toxicity of misonidazole alone in CHO cells is concentration dependent; it shows minimal toxicity to these cells at 5 mM while at 15 mM a 3-h exposure inactivates 95% of the cells. When 5 mM ascorbate is added to either 5 mM or 15 mM misonidazole there is a substantial increase in cell inactivation, although ascorbate alone is nontoxic to hypoxic cells. The

combined treatment of 5 mM ascorbate and 5 mM misonidazole is considerably more toxic than 15 mM misonidazole alone.

In contrast to its effect on hypoxic cells, ascorbic acid is very toxic to aerobic cells^{14,15}. This effect in CHO cells is shown in Fig. 2. Ascorbate oxidation involves one-electron transfer reactions which apparently result in the formation of hydrogen peroxide (H_2O_2) (ref. 16). The addition of catalase, an enzyme which blocks the accumulation of H_2O_2 , completely eliminates the aerobic toxicity of ascorbate to CHO cells. The addition of a nonspecific protein, bovine serum albumin, did not reduce the toxicity (data not shown).

Since ascorbate is an effective reducing agent, the increased toxicity of the combined misonidazole-ascorbate treatment of hypoxic cells could be due to accelerated formation of toxic, reduced metabolites of misonidazole. Alternatively, it might be argued that under hypoxia, misonidazole acts as an oxidant of ascorbate, leading to the formation of H_2O_2 and its resultant toxicity. But, catalase does not reduce the hypoxic toxicity of the misonidazole-ascorbate combination (Fig. 2). Furthermore, when gulonolactone, a nonreducing analogue of ascorbate, was substituted for ascorbate, there was no enhancement of misonidazole toxicity. It seems likely, therefore, that ascorbate promotes the reduction of misonidazole, leading to a greater accumulation of toxic derivatives. Another possible explanation is that, although ascorbate itself is not toxic to hypoxic cells it could somehow be predisposing cells to misonidazole damage. This possibility has not yet been examined.

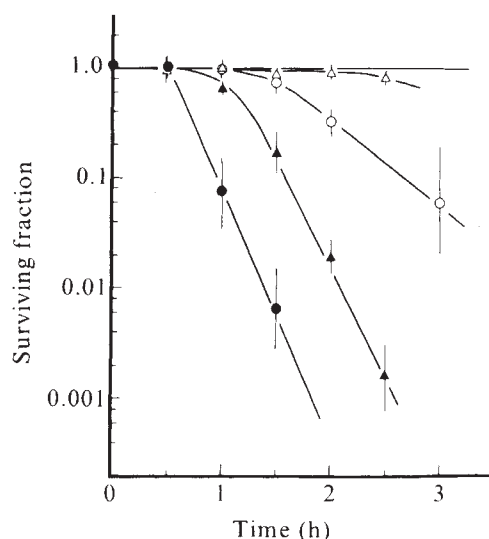


Fig. 1 Exponentially growing CHO cells were collected from suspension culture by centrifugation and resuspended in growth medium (MEM Alpha, supplemented with 10% foetal calf serum) at 6×10^6 cells ml^{-1} . After maintaining them at this concentration for 10–15 min at room temperature they were added to pre-gassed glass vessels containing growth medium and a prescribed amount of misonidazole and/or ascorbic acid such that the final concentration of cells was 2×10^5 cells ml^{-1} . Drug-medium solutions were freshly prepared before each experiment. The pH was adjusted to approximately 7.3 and solutions were sterile filtered. Purified N_2 , containing less than 5 p.p.m. O_2 , was flowed for 15 min over the stirred solutions at 37 °C before the addition of 1 ml of cells. The flow of N_2 continued throughout the exposure time. At the indicated times 1 ml samples were withdrawn and immediately diluted 10 times in aerobic growth medium at 0 °C. The cells were then washed once and plated for colony formation. Δ , 5 mM misonidazole; $\circ$, 15 mM misonidazole; $\blacktriangle$, 5 mM misonidazole plus 5 mM ascorbate; $\bullet$, 15 mM misonidazole plus 5 mM ascorbate. The error bars represent the s.e.m. values derived from pooled experiments.

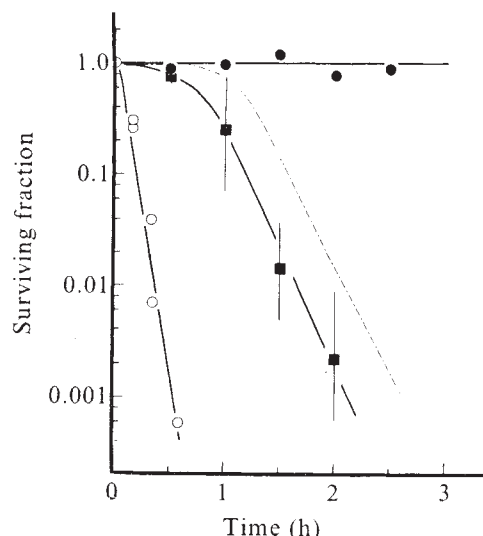


Fig. 2 The same procedure was used as described in Fig. 1, except that O_2 was flowed where indicated instead of purified N_2 . $\circ$, 5 mM ascorbate, O_2 ; $\bullet$, 5 mM ascorbate plus 0.3 mg ml^{-1} catalase, O_2 ; $\blacksquare$, 5 mM misonidazole plus 5 mM ascorbate and 0.3 mg ml^{-1} of catalase, N_2 . The dotted line represents the 5 mM misonidazole plus 5 mM ascorbate response in N_2 from Fig. 1. Under N_2 , the presence of catalase (or bovine serum albumin) increased slightly the toxic effects of the misonidazole-ascorbate combination in each experiment.

Several other aspects of this effect of ascorbate have been studied. In order to determine whether the supposed reaction between misonidazole and ascorbate required the presence of cells, ascorbate and misonidazole were added to the growth medium for up to 2 h before addition of cells. This did not change the combined response of Fig. 1, suggesting that the effect is a cell-mediated reaction, and does not occur spontaneously. The enhanced toxicity was found to be strongly temperature dependent. It is more rapid at 42.5 °C and almost absent at 0 °C. Variation of ascorbate concentration at a fixed misonidazole concentration, changes the duration of the initial shoulder seen in Fig. 1, but has little effect on the slope of the exponential portion of the curve. A similar potentiation of misonidazole toxicity was observed in the presence of reduced glutathione, the cofactor in the enzymatic oxidation of ascorbate; however, glutathione was less effective on a molar basis than ascorbate.

The effect of ascorbate and other reducing species (such as glutathione, cysteine, cysteamine) on the hypoxic cytotoxicity of misonidazole clearly requires further examination in order to assess and regulate any possible role in clinical applications. Misonidazole doses used in current clinical protocols result in plasma and tissue concentrations approaching 1 mM¹. The concentration of ascorbate (and perhaps other reducing agents) in the human body can also reach this range¹⁷, where the above *in vitro* studies suggest the occurrence of significant synergistic interactions. Whether the net effect of such interactions might be primarily positive (a contribution to the chemotherapeutic properties of misonidazole) or negative (an enhancement of the drug's cytotoxicity) cannot be predicted. The possibility, however, that such interactions may contribute to the problems of peripheral neuropathy, which have been described in reports of clinical trials of the drug¹, needs to be explored. *In vivo* studies in experimental animals are a logical next step.

We thank Mrs Isabel Harrison and Mr Spencer Kong for technical assistance. The misonidazole was supplied by Dr C. E. Smithen of Roche Products Ltd. This work was

supported by the BC Cancer Foundation and the National Cancer Institute of Canada. P.D.J. is the holder of a NRC postgraduate scholarship.

P. DAVID JOSEPHY
BRANKO PALCIC
LLOYD D. SKARSGARD

BC Cancer Foundation,
2656 Heather Street,
Vancouver, BC,
V5Z 3J3, Canada

University of British Columbia,
Vancouver, BC,
V6T 1W5, Canada

Received 24 October; accepted 17 November 1977.

- Gray, A. J., Dische, S., Adams, G. E., Flockhart, I. R. & Foster, J. L. *Clin. Radiol.* 27, 151-157 (1976).
- Dische, S., Gray, A. J. & Zanelli, G. D. *Clin. Radiol.* 27, 159-166 (1976).
- Thomlinson, R. H., Dische, S., Gray, A. J. & Errington, L. M. *Clin. Radiol.* 27, 167-174 (1976).
- Urtasun, R. C., Band, P. & Chapman, J. D. *Radiat. Res.* 70, 704 (1977).
- Moore, B. A., Palcic, B. & Skarsgard, L. D. *Radiat. Res.* 67, 459-473 (1976).
- Hall, E. J. & Roizin-Towle, L. *Radiology* 117, 453-457 (1975).
- Varghese, A. J., Gulyas, S. & Mohindra, J. K. *Cancer Res.* 36, 3761-3765 (1976).
- Whitmore, G. F., Gulyas, S. & Varghese, A. J. *Radiat. Res.* 70, 703 (1977).
- Stratford, I. J. & Adams, G. E. *Br. J. Cancer* 35, 307-313 (1977).
- Hall, E. J. & Biaglow, J. *Int. J. Radiat. Oncol. Biol. Phys.* 2, 521-530 (1977).
- Taylor, Y. & Rauth, A. M. *Radiat. Res.* 70, 702 (1977).
- Palcic, B. & Skarsgard, L. D. *Br. J. Cancer* (submitted).
- Hall, E. J., Astor, M., Geard, C. & Biaglow, J. *Br. J. Cancer* 35, 809-815 (1977).
- Benade, L., Howard, T. & Burk, D. *Oncology* 23, 33-43 (1969).
- Peterkofsky, B. & Prather, W. J. *Cell. Physiol.* 90, 61-70 (1977).
- Mapson, L. W. in *The Vitamins* (eds Sebrell, Jr, W. H. & Harris, R. S.) 2nd edn, 386-398 (Academic, New York, 1967).
- Lewin, S. in *Vitamin C: Its Molecular Biology and Medical Potential* 78-79 (Academic, London 1976).

Novel treatment for joint inflammation

THE use of liposomes in the entrapment and administration of several therapeutic agents has been described for the treatment of cancer¹ and heavy metal poisoning², and in enzyme replacement therapy³. The advantages of a liposome drug over the use of free or polymer-bound agent are decreased toxicity and degradation, the use of smaller doses and the possibility of targeting the liposome towards a given tissue or site⁴. The localised administration of therapeutic agents, such as the intra-articular injection of cortisol esters in the treatment of rheumatoid arthritis⁵, is a situation where the use of liposomes as a means of providing a stable particulate suspension of entrapped steroid might be used to advantage. This method should reduce considerably the effective dose required to produce relief, and diminish side effects due to escape of steroid from the joint⁶. We report here that the treatment of an acute experimental arthritis in the rabbit with a liposome preparation containing cortisol results in a significant reduction in joint temperature and diameter, whereas treatment with an equivalent amount of cortisol alone, or with liposomes lacking cortisol, does not reduce these two parameters of inflammation.

Old English rabbits (body weight 1.5-2.0 kg) were used throughout the study. A bilateral inflammation of the knee joints was induced by the intra-articular injection of a complex formed from 7.5 mg poly-D-lysine (molecular weight 150,000) and 7.5 mg hyaluronic acid⁷ (both Miles). The initial inflammation induced by the complex is an acute synovitis with fibrin exudation and polymorphonuclear leukocyte infiltration. After 1 week this stage merges with a subacute phase showing macrophage infiltration. At 2-3 weeks after induction the arthritis is fully developed with lymphocytic aggregations, plasma cell infiltration, and the appearance of pannus. Evidence of active chronic inflammation then persists for 3-4 months, after which signs of increased fibrotic scarring occur. This process is normally complete within 7 months.

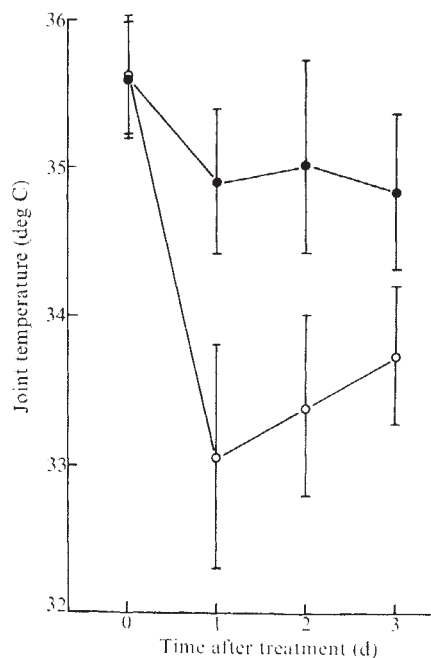
Infra-red radiation from the surface of the joint and its maximum diameter were used to follow the course of inflammation and to monitor the effect of treatment. The liposome

preparation containing cortisol palmitate was prepared as previously described⁸ using dipalmitoylphosphatidylcholine and egg yolk phosphatidic acid. Control liposome preparations were made by omitting the cortisol palmitate. Control liposomes, liposomes containing cortisol palmitate, or cortisol acetate particles in saline suspension were injected intra-articularly into one joint 4 d after induction of the inflammation. The effect of treatment on joint temperature and diameter was monitored for a period of 3 to 6 d.

Treatment of the experimental arthritis with a liposome preparation containing 35.5 µg cortisol palmitate (equivalent to 20 µg cortisol) resulted in a significant and sustained fall in the temperature (Fig. 1) and the diameter (Fig. 2) of the injected joint. The reduction in both these parameters of inflammation occurred within 24 h and there was no evidence of a contralateral effect in the uninjected joint. Neither injection of a liposome preparation containing 35.5 µg of the palmitate ester of 11- α -cortisol, a stereoisomer of cortisol which has no glucocorticoid activity⁹, nor a control liposome preparation containing no steroid had any effect on temperature or diameter. The injection of control liposomes, or liposomes containing cortisol palmitate into normal rabbit joints did not cause an increase in temperature or diameter, and it was concluded that the preparations were devoid of any inflammatory activity. There have been previous reports of liposomes having biological activity in their own right¹⁰, but the liposomes used in this study seemed to be without activity. The injection of cortisol acetate (equivalent to 20 µg cortisol) in saline suspension did not reduce the temperature or diameter of inflamed joints. Anti-inflammatory activity was seen at much higher doses (equivalent to 0.2-2.0 mg cortisol) and was restricted to a 24 h fall in temperature and diameter, whereas liposomes containing cortisol palmitate produced a sustained (3-5 d) reduction in these parameters of inflammation.

It was shown previously⁸ that cortisol esters are retained in liposomes prepared from dipalmitoylphosphatidylcholine and phosphatidic acid only if there is a hydrophobic anchor.

Fig. 1 Effect of liposomes containing cortisol palmitate on joint temperature. Hair was removed from the joint before the start of the experiment with a commercial depilatory cream. The surface temperature of an area 6 mm in diameter situated laterally over the joint space was measured using a Heimann KT 41 radiation thermometer. The joint surface temperature was allowed to equilibrate with the controlled ambient temperature ($20.5 \pm 0.25^\circ\text{C}$) before measurements were taken. Values are mean $\pm$ s.e.m. of 6 animals. ○, Treated joint; ●, untreated joint.



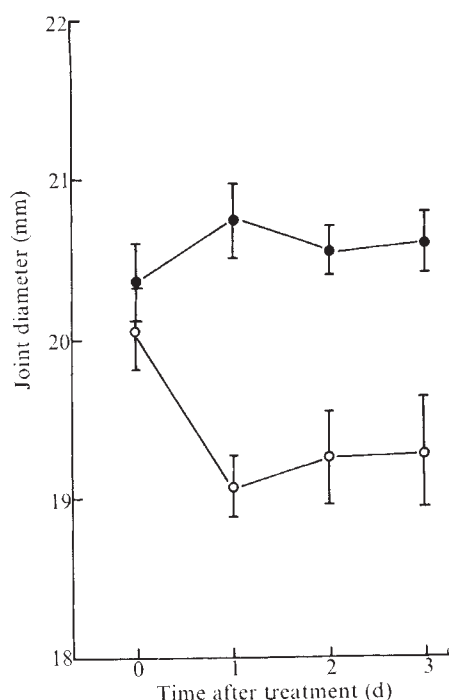


Fig. 2 Effect of liposomes containing cortisol palmitate on joint diameter. Maximum joint diameter was measured using a Baly spring-loaded micrometer. Values are mean $\pm$ s.e.m. of 6 animals. $\circ$, Treated joint; $\bullet$, untreated joint.

The demonstration that an injection of cortisol palmitate incorporated into a liposome preparation has a much greater anti-inflammatory activity than an equivalent dose of microcrystalline cortisol acetate, a standard anti-inflammatory steroid ester, validates the use of liposome preparations for local delivery of steroids.

This work was carried out at Strangeways Research Laboratory, Cambridge, and at ICI Pharmaceuticals Division, Alderley Park, Cheshire. Similar results were obtained at both centres.

J. T. DINGLE
J. L. GORDON
B. L. HAZLEMAN
C. G. KNIGHT
D. P. PAGE THOMAS
N. C. PHILLIPS
I. H. SHAW

Strangeways Research Laboratory,
Wort's Causeway,
Cambridge, UK

F. J. T. FILDES
JEAN E. OLIVER
G. JONES
E. H. TURNER
J. S. LOWE

Imperial Chemical Industries Limited,
Pharmaceuticals Division,
Mereside,
Alderley Park,
Macclesfield,
Cheshire, UK

- Rahman, Y. E., Rosenthal, M. W., Cerny, E. A. & Moretti, E. S. *J. Lab. clin. Med.* 83, 640-647 (1974).
- Gregoriadis, G., Leathwood, P. D. & Ryman, B. *FEBS Lett.* 14, 95-99 (1971).
- Fendler, J. H. & Romero, A. *Life Sci.* 20, 1109-1120 (1977).
- Hollander, J. L. in *Arthritis and Allied Conditions* (eds Hollander, J. L. & McCarthy, D. J.) 8th edn, 517-534 (Lea and Febiger, Philadelphia, 1972).
- Hess, E. V. & Coldman, J. A. in *Arthritis and Allied Conditions* (eds Hollander, J. L. & McCarthy, D. J.) 8th edn, 506-512 (Lea and Febiger, Philadelphia, 1972).
- Page Thomas, D. P. in *Bayer-Symposium VI, Experimental Models of Chronic Inflammatory Diseases* (eds Glynn, L. E. & Schlumberger, H. D.) 353-365 (Springer, Berlin, 1977).
- Shaw, I. H., Knight, C. G. & Dingle, J. T. *Biochem. J.* 158, 473-476 (1976).
- Simonsson, B. *Steroids* 20, 23-39 (1972).
- Bruni, A., Tottano, G., Leon, A. & Boararo, E. *Nature* 260, 331-333 (1976).

Relationship between haptoglobin and *Streptococcus pyogenes* T4 antigens

DURING a survey of agglutinating activity of human saliva against streptococci¹ we found that saliva from individuals of blood group O agglutinated a particular strain of group G streptococci more often than did serum from those with blood group A or B. When we tested 833 sera we found a dimorphism for antibody-like substances against group G streptococci—sera could be divided into two groups according to their agglutination titre against group G streptococci. The first group had low titres, not rising above 1:4 and more usually 1:1; the second group had much higher titres ranging from 1:200 to 1:3,200. There was no overlap between the two groups of sera. We thought that this might be related to a genetic system, probably diallelic. We now report a relationship between the haptoglobin genotype of serum and the streptococcal antigens with which it reacts.

We tested our samples for a wide range of polymorphic blood groups (ABO, MNSS, P₁/P₂, rhesus, K/k, Le^a/Le^b, Duffy and Kidd), serum groups (Gm 1, Gm 2, Gm 4, Hp, Gc), enzyme groups (acP, PGM, ADA, AK, GPT and EsD) and HLA groups (factors of loci Sd¹ and Sd²). All sera without agglutinating activity or with low titres against a particular strain of group G streptococci were of the haptoglobin (Hp) type Hp 1-1. All sera with high titres, however, were of Hp types 2-1 or 2-2. There was no effect due to sex or age.

Further studies showed that sera with the Hp² gene product reacted with *Streptococcus pyogenes* strains carrying the T antigen complex T4/24. This antigen may also be present in some C and G group streptococci and it was only these strains that were agglutinated with high titres by Hp 2-2 and Hp 2-1 sera. In our series 1.4 % of group C streptococci and 6.4 % of group G streptococci carried the T4/24 antigen. A relationship between the T4/24 antigen of streptococci and haptoglobin types has not been described before.

The mechanism of the reaction is unknown but the reacting substance in all human sera is haptoglobin itself. This was demonstrated in cooperation with Dr Uhlenbruck in Cologne, using a purified haptoglobin preparation (98 % purity, Behringwerke, Marburg) which agglutinated the streptococci in the same way. Further evidence was given by a survey of human cord blood samples. If agglutination was negative, haptoglobin could not be detected. Animal sera with Hp 1-1-like patterns (rhesus monkey, rabbit, pig and sheep) did not agglutinate the T4/24 streptococci or did so only to low titres.

Finally, we found that the agglutinating activity of Hp 2-2 or Hp 2-1 sera could be inhibited by the addition of Hp 1-1 serum of human or animal origin before the test. We believe that Hp 1-1 protein represents a form of 'blocking antibody' for the appropriate streptococci.

W. KÖHLER
O. PROKOP

Academy of Science of the DDR,
Jena, DDR

Received 1 July; accepted 25 November 1977.

1. Prokop, O. & Köhler, W. *Z. Immun. Forsch.* 152, 362-365 (1977).

Received 30 September; accepted 30 November 1977.

I. Gregoriadis, G., Swain, C. P., Willis, E. J. & Tavill, A. S. *Lancet* i, 1313-1314 (1974).

Partial $\gamma 1$ fragment is a 'natural' bovine IgG1 fragment without detectable class-specific γ determinants

IgG in bovine sera generally consists of a mixture of the two subclasses IgG1 and IgG2. While testing the reliability of the radial immunodiffusion (RID) method for quantifying IgG in bovine body fluids¹, we found a discrepancy. With most sera tested, the combined amounts of IgG1 and IgG2 determined using subclass-specific antisera was equal to the amount of IgG determined with class-specific antisera against γ determinants, that is common to IgG1 and IgG2. But with a few sera the amounts were not equal (unpublished results). (Class-specific antisera against γ determinants were produced by removing antibodies against subclass determinants on immunoabsorbent columns, starting from antisera which contained antibodies against both class ($=\gamma$) and subclass ($=\gamma 1$ and/or $\gamma 2$) determinants.) Using immunoelectrophoretic (IE) analysis, we have now shown that those sera contain a 'fast' IgG component as well as both IgG subclasses. The third component was represented by an arc in the anodic region of the precipitation pattern, and it proved partially identical to that of IgG1 (Fig. 1). Such an arc can be seen in a photograph of IE patterns of sera from calves infected with *Trypanosoma vivax*, although the arc is not mentioned in the report².

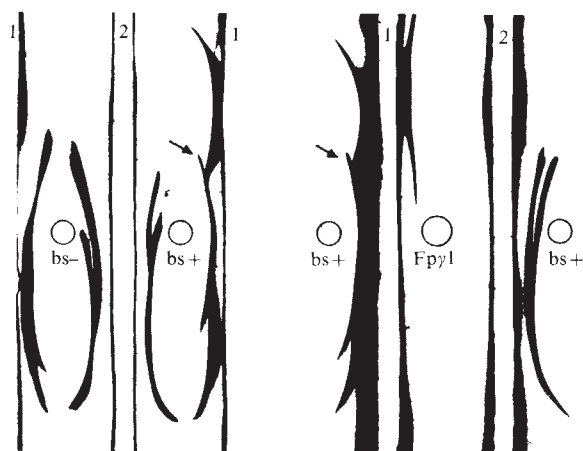


Fig. 1 bs+, bovine serum containing Fp $\gamma 1$ (6 μ l); bs-, normal bovine serum (6 μ l); Fp $\gamma 1$, partial $\gamma 1$ fragment (24 μ l, 1 mg ml⁻¹); 1, goat anti-bovine IgG1; 2, goat anti-bovine IgG2. The anode is at the top. Arrows indicate reaction of partial identity of Fp $\gamma 1$ and IgG1

The arc in Fig. 1 was produced with antisera against IgG1 that contained antibodies against γ and $\gamma 1$ determinants. When we used antiserum against IgG2 that contained antibodies against L chain, $\gamma 2$ and γ determinants, the arc was absent (Fig. 1). The absence of $\gamma 1$ determinants and the presence of $\gamma 1$ determinants on the proteins in this arc explained the discrepancy in RID values. The IE arc seemed to represent all or a part of a special kind of IgG1. Therefore, and for reasons explained below, this protein is referred to as partial $\gamma 1$ fragment (Fp $\gamma 1$).

Serum samples containing Fp $\gamma 1$ were collected monthly from cows in connection with an investigation into the occurrence of parasitic infections in the field³. They were stored at -20 °C. It was unlikely that 'ageing' of sera gave rise to Fp $\gamma 1$ because some freshly collected serum contained Fp $\gamma 1$ whereas older serum from the same animals did not. Sometimes serum collected from the same animal in two successive months produced the IE arc, whereas samples collected shortly before and after this period contained little or no Fp $\gamma 1$.

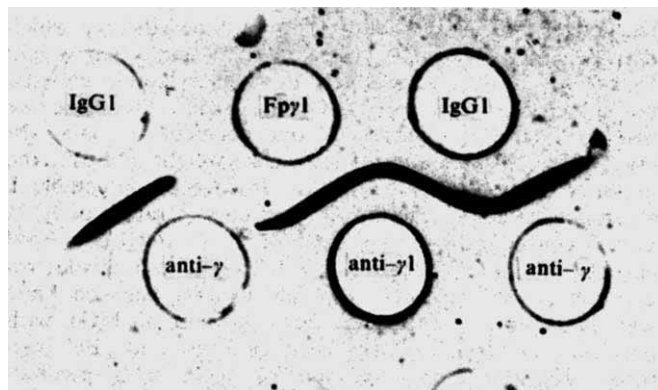
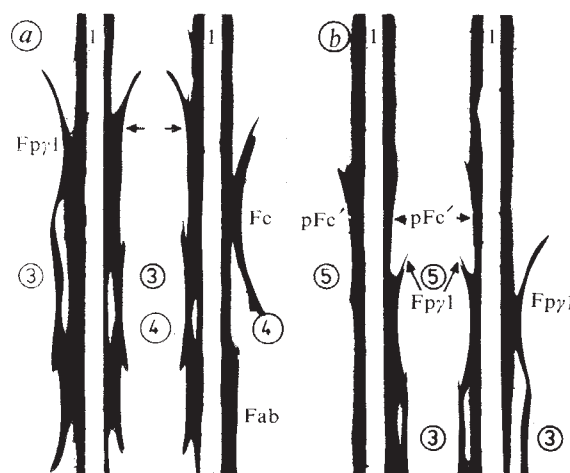


Fig. 2 The concentration of IgG1 was ± 0.1 mg ml⁻¹, and of Fp $\gamma 1$ ± 1.0 mg ml⁻¹, anti- γ , goat antiserum specific for IgG class determinants; anti- $\gamma 1$, goat antiserum specific for IgG1 subclass determinants. Note the reaction of complete identity of IgG1 and Fp $\gamma 1$ in the reaction with antiserum specific for $\gamma 1$ determinants and the absence of the precipitation reaction of Fp $\gamma 1$ with antiserum specific for γ determinants.

To characterise Fp $\gamma 1$, we isolated it on immunoabsorbent columns⁴ loaded with purified antibodies against $\gamma 1$ determinants (column 1) and γ determinants (column 2). Serum samples containing Fp $\gamma 1$ were applied to column 1 and the mixture of Fp $\gamma 1$ and IgG1 was eluted with 3M ammonium thiocyanate in phosphate-buffered saline. When this mixture was applied to column 2, only IgG1 was bound.

The protein thus isolated produced an IE arc comparable with the Fp $\gamma 1$ arc from the original serum (Fig. 1). In double immunodiffusion (DID), with antiserum specific for $\gamma 1$ determinants Fp $\gamma 1$ gave a reaction of complete identity to IgG1 but did not react with four class-specific antisera specific for γ determinants, although these reacted with IgG1. The reaction with one of these antisera is shown in Fig. 2. The isoelectric point of Fp $\gamma 1$, determined by isoelectric focusing⁵, was between 4.4 and 4.7. Estimates of the molecular weight of Fp $\gamma 1$, obtained by gel-filtration using Sephadex G-200 Fine⁶ and electrophoresis in polyacrylamide gel⁷ without reducing agents were 42,000 and 44,000

Fig. 3 Double immunoelectrophoresis patterns of Fp $\gamma 1$ with Fc fragment (a) and pFc' fragment (b). The anode is at the top. Note the absence of Fc spurs at the inner side of the Fp $\gamma 1$ arcs (arrows in a) and the presence of Fp $\gamma 1$ spurs at the inner side of the pFc' arcs, indicating reactions of complete and partial identity, respectively. The splitting of the Fp $\gamma 1$ arcs at the inner side of pFc' arcs is brought about by the presence of the pFc' arcs, preventing antibodies against determinants common to pFc' and Fp $\gamma 1$ from reaching Fp $\gamma 1$. 1, Goat antiserum against IgG1; 3, bovine serum containing Fp $\gamma 1$; 4, Fc/Fab mixture; 5, pFc'.



respectively. Reduction with 0.1% dithioerythritol, which disassociates the H and L chains of IgG1 (molecular weight 58,000 and 23,500, respectively in polyacrylamide gel) did not affect the value of 44,000 for Fpy1. Therefore we assumed that Fpy1 consists of one polypeptide chain. On the basis of the estimates of molecular weight of Fpy1; the presence of $\gamma 1$ determinants; the absence of detectable L chain determinants, and a decreased concentration of IgG1 in sera containing Fpy1 (see below), we assumed that Fpy1 emanated from an H chain of IgG1. Double immunoelectrophoretic analysis⁸ showed complete identity between Fpy1 and Fc fragment (obtained after digestion of IgG1 with papain⁹) and partial identity between Fpy1 and pFc' fragment (obtained after digestion of IgG1 with pepsin¹⁰) (Fig. 3).

The pFc' fragment includes the complete COOH-terminal part of the H chain (CH3 region) and the Fc fragment approximately covers the combined CH2 and CH3 regions. The molecular weight of Fpy1 (42,000–44,000) accounts for about three-quarters of the estimated molecular weight of an H chain. This molecular weight (58,000) is in fair agreement with estimates of molecular weights of bovine IgG1 H chains reported before (for example, ref. 11). Although the precise location of Fpy1 in the H chain remains to be determined (a small fragment may be missing from one or both sides of the proposed polypeptide chain) Fpy1 probably consists of the combined CH1, CH2 and CH3

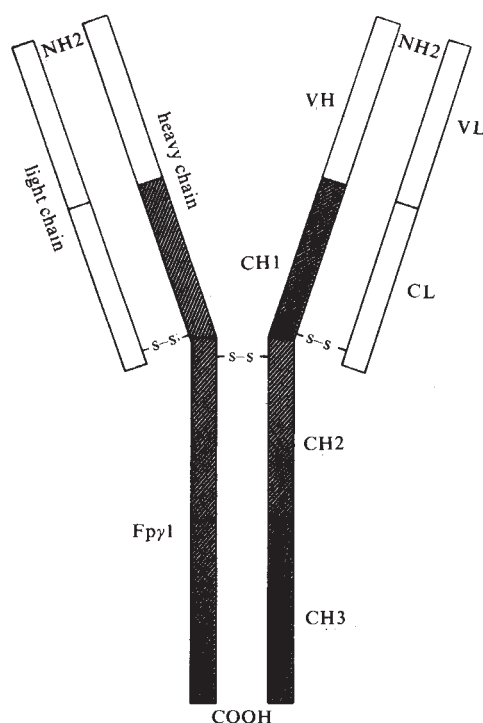


Fig. 4 Schematic drawing of the linear structure of IgG1, with an indication of variable (V) and constant (C) regions and the probable position of the Fpy1-fragment (hatched parts).

regions of the heavy chains of the IgG1 molecule (Fig. 4). One may argue instead that part or all of the VH region forms part of Fpy1, so that there are deletions in the H chains as in heavy-chain disease proteins¹². This is unlikely in view of the presence of the CH2 and CH3 regions in Fpy1 and because sera collected shortly before and during the appearance of Fpy1 always contained IgG1 with H chains of normal molecular weight. We assume that this IgG1 is the source of Fpy1, for the concentration of IgG1 is strongly reduced in sera containing Fpy1.

It is surprising that class-specific γ determinants cannot be detected on Fpy1 with either precipitation techniques (IE and DID) or immunoadsorbents. Perhaps γ determinants are more labile than $\gamma 1$ determinants, or they are on the Fd part of IgG1 nearer to the VH/CH1 boundary than $\gamma 1$ determinants, or near interchain disulphide bonds.

We have not detected Fpy1 in hundreds of normal sera, in spite of a supposed continuous catabolism of antibodies directed against various antigens. But Fpy1 appears in sera of cows intentionally infected with *Fasciola hepatica*, and it is likely that the sera in which Fpy1 was found originally were from similarly infected cows. Our results suggest that while antibodies are being bound to antigens of parasitic origin, combined CH1, CH2 and CH3 regions of $\gamma 1$ chains, and presumably CL regions as well, are separated from the rest of the IgG1 molecule, breaking disulphide bonds. Whether CL regions remain intact for some time and appear in Fpy1-positive serum is not known.

The chance of detecting proteins consisting of CL regions only is small in view of the short half-life of light chains¹³. We assume that Fpy1 is generated while the immunological system is confronted with a large antigenic load, leading to, for example, incomplete phagocytosis of the antigen-antibody complexes. It remains to be seen whether Fpy1 arises only in response to parasitic or other corpuscular antigens or whether large amounts of any antigen lead to comparable fragmentation of IgG1. Although fragmentation has been described for Fc- and Fab-like fragments in normal urine¹⁴ and plasma¹⁵, Fpy1 seems to be the first reported catabolic fragment of Ig that is probably related to antigen-antibody interactions *in vivo*.

I thank Mrs Jacqueline J. L. Noordijk and Mrs Willy Stok for technical assistance and Dr L. J. van Tiggele for bovine sera.

B. A. BOKHOUT

Central Veterinary Institute,
PO Box 6007,
3002 AA Rotterdam, The Netherlands

Received 26 September; accepted 25 November 1977.

1. Bokhout, B. A. *J. Immun. Meth.* 7, 187–198 (1975); 199–210 (1975).
2. Clarkson, M. J., Penhale, W. J. & McKenna, R. B. *J. comp. Path.* 85, 397–410 (1975).
3. Tiggele, L. J. van & Over, H. *J. Vet. Path.* 1, 239–248 (1976).
4. Bokhout, B. A. & van Tiggele, L. *J. Sci. Tools* 24, 56–57 (1977).
5. LKB Application Note 2117, 1 (1973).
6. Andrews, P. *Biochem. J.* 91, 222–233 (1964).
7. Weber, K. & Osborn, M. *J. biol. Chem.* 244, 4406–4412 (1969).
8. Bokhout, B. A., *J. Immun. Meth.* 5, 229–238 (1974).
9. Fleishman, J. B., Pain, R. H. & Porter, R. R. *Arch. biochem. biophys.*, Suppl. 1, 174–180 (1962).
10. Leslie, R. G. Q., Melamed, M. M. & Cohen, S. *Biochem. J.* 121, 829–837 (1971).
11. Brock, J. H., Arzabe, F. R., Ortega, F. & Pineiro, A. *Immunology* 32, 215–219 (1977).
12. Natvig, J. B. & Kunkel, H. G. *Adv. Immun.* 16, 1–59 (1973).
13. Spiegelberg, H. L. & Weigle, W. O. *J. Immun.* 95, 1034–1040 (1966).
14. Turner, M. W. & Rowe, D. S. *Nature* 210, 130–132 (1966).
15. Berggard, I. & Bennich, H. *Nature* 214, 697–699 (1967).

Production of HLA antibody in mice immunised with syngeneic mouse-human hybrid cells containing human chromosome 6

XENOIMMUNISATION with human cells or isolated molecules found on the surface of human cells results in the production of antibodies to a variety of human alloantigens (refs 1–3 and unpublished results). The genetic loci which specify the antigens being recognised can then be determined by using rodent-human hybrids containing various human chromosomes as target cells. In this manner, chromosome 11 has been found to include genes coding for a series of antigens designated A_L (refs 1, 4) or HSA-1 (ref. 2) and chromosome 15 to code for β_2 microglobulin⁵. Murine-human somatic cell hybrids with one or a limited number of human chromosomes have been used by us and by others to obtain antisera in mice syngeneic to the mouse parental cells reactive with antigens coded for by

specific human chromosomes^{2,5-7}. The advantage of this method is that antisera generated are specific for the antigen(s) coded for by a particular human chromosome and the reactivity is to a limited number of human cell surface antigens. This allows for both systematic genetic mapping of these antigens and detection of previously unknown cell surface molecules. Chromosome 6 has been shown both by pedigree analysis⁸ and by the use of interspecific somatic cell hybrid target cells⁹ to be the location of the genes coding for the serologically defined antigens (HLA) of the major histocompatibility complex (MHC). We have used a human-mouse somatic cell hybrid containing several human chromosomes, including chromosome 6, to immunise mice syngeneic to the mouse parental cell. The reactivity of the absorbed antiserum has been analysed both serologically and by sodium dodecyl sulphate polyacrylamide gel electrophoresis after immunoprecipitation. Evidence is presented here to show that mice immunised with somatic cell hybrids containing human chromosome 6 produce antibody which reacts with HLA as well as with other human lymphocyte cell surface antigens.

Hybrid cells used for immunisation were the Nude 1 (N 1) cell line¹⁰. This hybrid line was derived from a tumour grown in an athymic mouse injected with hybrid cells obtained by the Sendai virus-mediated fusion of C57B1/6 mouse peritoneal macrophages and LN-SV (an SV40-transformed human cell line derived from skin fibroblasts of a Lesch-Nyhan syndrome patient)¹¹. N 1 was found by isozyme analysis and karyotypic evaluation to contain human chromosomes 6, 7, 14 and 21. Clones containing chromosomes 7, 14 and 21 but not 6 have also been derived from this original hybrid.

The HLA type of human parental cell LN-SV was determined by the standard NIH microcytotoxicity assay in Terasaki plates. These cells were found to express HLA-A2 and HLA-B5 antigens. Clonally derived hybrid cells containing one copy of human chromosome 6 were used to absorb several cytotoxic HLA typing sera, specific for HLA-A2 and HLA-B5. These sera were then tested on a panel of human lymphocytes of known HLA type. Table 1 shows that the hybrid cells containing human chromosome 6 expressed HLA-A2, and were HLA-B5-negative. Hybrid cells without human chromosome 6 (human chromosomes 7, 14, and 21 present) did not absorb out HLA reactivity.

To determine if the murine humoral response to somatic cell hybrids containing human chromosome 6 is to MHC-defined antigens and/or to species-defined antigens, C57B1/6J mice were immunised with N 1 hybrid cells. At 7-10 d after each intraperitoneal injection of $1-3 \times 10^7$ trypsinised cells, mice were bled, and serum was separated and stored at -70°C until use. For these studies, anti-N 1 serum from the sixth immunisation was absorbed with another human-mouse

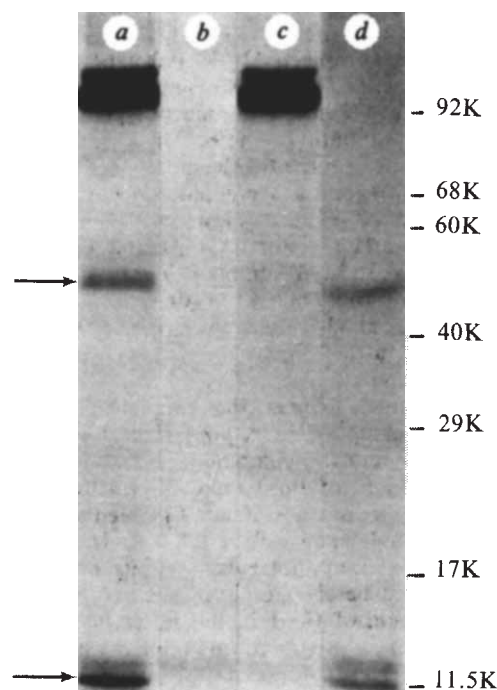


Fig. 1 Immunoprecipitation of ^{125}I -labelled surface proteins from human lymphocytes with absorbed anti-N 1 serum. Human peripheral blood lymphocytes (isolated from interface of Ficoll-Hypaque gradient (density 1.077) and washed three times with Dulbecco's modified PBS without Ca^{2+} and Mg^{2+} salts) were iodinated (New England Nuclear, NEZ-033H) by the lactoperoxidase-glucose oxidase method as previously described^{12,17}. Extracts were prepared by 1 h incubation (0°C) in 0.1% NP-40, 0.1 M Tris, pH 6.8, 15% glycerol and 2 mM phenyl-methylsulphonyl fluoride (PMSF). Iodinated extracts were incubated for 1 h, 0°C , at a 2:1 ratio with normal human serum from the donor of the lymphocytes or with rabbit anti-human β_2 microglobulin. Immune complexes were cleared from the samples by incubation at 0°C for 1 h with protein A-bearing *Staphylococcus aureus* Cowan 1 strain. Specific serum at a 1:5 final dilution was added to the precleared samples. After incubation at 0°C for 1 h, the immune complexes were absorbed to *S. aureus*, centrifuged and washed. After elution of the absorbed proteins from the bacterial surface (boiling *S. aureus* in 0.1 M Tris, pH 6.8, 15% glycerol, 2 mM PMSF, 2% sodium dodecyl sulphate (SDS) and 1 mM dithiothreitol) the bacteria were removed by centrifugation and the samples applied to 10% SDS-acrylamide gels with 5% stacking gels¹³. Gels were dried onto filter paper and exposed to Kodak NS2T X-ray film. Molecular weight markers for gel calibrations: cytochrome *c* (11,500), myoglobin (17,000), carbonic anhydrase (29,000), aldolase (40,000), catalase (60,000), bovine serum albumin (68,000), phosphorylase A (92,000); a, absorbed anti-N 1 serum, sample precleared with normal human serum; b, normal mouse serum, sample precleared with normal human serum; c, absorbed anti-N 1 serum, sample precleared with rabbit anti-human β_2 microglobulin; d, rabbit anti-human β_2 microglobulin serum, sample precleared with normal human serum.

Table 1 HLA expression on hybrid cells

HLA alloantisera specificity*	Absorbing hybrid	Human chromosome	HLA of lymphocytes A1, 2; B8	A2, 29; B5, 12
A2 ^a	C1 35	+	0†	0
A2 ^a	C1 37	—	4 ⁺	4 ⁺
A2 ^b	C1 35	+	0	0
A2 ^b	C1 37	—	4 ⁺	4 ⁺
B5 ^c	C1 35	+	NR‡	3 ⁺
B5 ^c	C1 37	—	NR	3 ⁺
A2B5 ^d	C1 35	+	0	4 ⁺
A2B5 ^d	C1 37	—	3 ⁺	4 ⁺
A1B8 ^e	C1 35	+	4 ⁺	NR
A1B8 ^e	C1 37	—	3 ⁺	NR

*Sera obtained from Transplantation Immunology Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health: a, 2-50-6-03-15-03; b, 1-01-7-02-07-02; c, 2-52-1-01-14-01; d, 2-58-9-02-21-04; e, 1-04-1-02-15-01.

†0-4⁺, strength of cytotoxic reaction; 0, 0-10%; 1⁺, 10-30%; 2⁺, 30-60%; 3⁺, 60-80%; 4⁺, 80-100%.

‡No reaction; lymphocytes do not express this antigen.

hybrid cell line of the same origin containing human chromosomes 7, 14 and 21. Testing revealed that the absorbed anti-N1 serum reacted only with hybrids containing chromosome 6.

To define the cell surface molecules with which the absorbed anti-N 1 serum reacted, we prepared extracts of human peripheral blood lymphocytes which had been enzymatically labelled with ^{125}I using the lactoperoxidase-glucose oxidase method¹². The extracts were reacted with absorbed anti-N 1 serum, or normal mouse serum, and the immune complexes were analysed on 10% polyacrylamide gels¹³. As shown in Fig. 1a, reaction of lysates with absorbed anti-N 1 serum results in precipitation of several proteins; two of these (see arrows) have molecular weights consistent with those of HLA and its associated light chain, β_2 microglobulin¹⁴. No radio-labelled bands are seen when normal mouse serum is used (Fig.

1b). When the extracts were pre-cleared by treatment with rabbit anti-human β_2 microglobulin serum (Dakoimmunoglobulin; Copenhagen, Denmark) (see Fig. 1 legend) and then reacted with anti-N 1 serum, the 47,000 molecular weight (MW) and 12,000 MW bands disappeared, leaving the higher molecular weight proteins coded for by human chromosome 6 (Fig. 1c). Electrophoresis of immune complexes of anti-human β_2 microglobulin serum with the labelled human lymphocytes shows β_2 microglobulin, the 12,000 MW band, and the co-precipitated HLA (Fig. 1d).

Further evidence that the reactivity to HLA is provided by lysostrip experiments shown in Table 2 (procedure and theory described in refs 15 and 16). Normal human lymphocytes of known HLA specificities after reaction with anti-N 1 are incubated with goat anti-mouse IgG. The cells are then reacted with HLA typing sera. The fact that reactivity to each HLA specificity is affected by this stripping procedure provides evidence that anti-N 1 serum is reacting with HLA molecules on the lymphocyte surface. It is important to note that although HLA-A2 is the only specificity shown to be present on the immunising cell surface, both HLA-A and HLA-B antigens are affected by the stripping procedure. This has been observed with antisera and lymphocytes of other HLA types (data not given). The response seems to be either to a common determinant of HLA molecules or a multiclonal response to many portions of the HLA molecule. If some B cell clones were responding to the private HLA specificities present on the immunising hybrid, they would not be detected because of the overwhelming response to other portions of the molecule. Immunising mice with hybrids containing human chromosome 6 and hybridising B cells from these mice to myeloma cells may result in the constitutive production of monoclonal antibodies specific for the HLA allele(s) of the immunising cell. *In vitro* xenoimmunisation of both human¹⁷ and rat¹⁸ cells with mouse lymphoid cells results in the development of cytotoxic effector cells which are specific for the immunising H2 haplotype. It seems that limited clonal stimulation occurs with the major reactivity limited to the recognition of the allelic specificities of H2.

Table 2 Susceptibility of human lymphocytes to cytotoxic killing after lysostripping with anti-N 1 and goat anti-mouse IgG treatment

Antiserum tested with*	% killed†
N1	7
HLA-A1 ^a	24
HLA-A2 ^b	1
HLA-A2 ^c	28
HLA-B8 ^d	32
HLA-B8 ^e	10
Human‡	100

2.5×10^5 Ficoll-Hypaque-separated human peripheral blood lymphocytes (HLA-A1, 2; B8) were treated with absorbed anti-N 1 serum (or normal mouse serum as control) 1:20, 30 min, 4 °C; treated with goat anti-mouse IgG, 1:5; 30 min, 4 °C, incubated 30 min., 37 °C; treated with goat anti-mouse IgG, 1:5, 30 min, 4 °C; washed three times between each treatment incubated 30 min, 37 °C, and washed twice. Volumes were kept constant at 100 μ l. All washes were with 1 mM HEPES-buffered RPMI 1640 containing 10% foetal bovine serum. Antisera were diluted in 1 mM HEPES- buffered RPMI 1640. This procedure was to 'induce' resistance to antibody-mediated complement-dependent cytotoxicity for the antigen(s) recognised by the anti-N 1 serum. The 'lysostripped' or treated control cells were then tested in an antibody-dependent complement-mediated microcytotoxicity assay to determine whether the cells were resistant to lysis by the indicated sera. (For procedure and discussion of theory, see refs 15 and 16.)

*Sera from Transplantation Immunology Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health: a, 2-65-9-11-17-08; b, 2-50-6-03-15-03; c, 1-01-7-02-07-02; d, 2-50-6-10-04-01; e, 2-51-2-04-27-01.

†Calculated as follows:

$$\left(\frac{\% \text{ anti-N 1 treated dead} - \% \text{ c' control dead}}{\% \text{ normal serum treated dead} - \% \text{ c' control dead}} \right) \times 100$$

The test sera used after lysostrip treatment were used at a dilution that killed > 60% of the normal serum control; in most cases 80–90% of control cells were killed at the dilution selected.

‡Serum from C57B1/6 mice immunised with LN-SV human cells.

We suggest that human-mouse hybrid cells containing human chromosome 6 can be used to immunise mice to produce antibodies to the major histocompatibility complex. Additional cell surface molecules coded for by human chromosome 6 are detected with the anti-N 1 sera.

This work was supported, in part, by a Career Development Award from the US NCI and grants from the NCI, American Cancer Society, National Foundation-March of Dimes, and a NSF Fellowship (R.M.).

DAVID P. ADEN
RACHEL MAUSNER
BARBARA B. KNOWLES

*The Wistar Institute of Anatomy and Biology,
36th Street at Spruce,
Philadelphia, Pennsylvania 19104*

Received 11 October; accepted 21 November 1977.

1. Puck, T. T., Wuthier, P., Jones, C. & Kao, F.-T. *Proc. natn. Acad. Sci. U.S.A.* **68**, 3102–3106 (1971).
2. Buck, D. W. & Bodmer, W. F. *Cytogenet. Cell Genet.* **14**, 257–259 (1975).
3. Goodfellow, P. N. *et al. Nature* **254**, 267–269 (1975).
4. Moore, E. E., Jones, C. & Puck, T. T. *Cytogenet. Cell Genet.* **17**, 89–97 (1976).
5. Buck, D. W., Goss, S. J. & Bodmer, W. F. *Cytogenet. Cell Genet.* **16**, 99–100 (1976).
6. Revel, M., Bach, D. & Ruddle, F. H. *Nature* **260**, 139–141 (1976).
7. Aden, D. P. & Knowles, B. B. *Immunogenetics* **3**, 209–221 (1976).
8. Lamm, L. U. *et al. Hum. Heredity* **24**, 273–284 (1974).
9. Van Someren, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **71**, 962–965 (1974).
10. Croce, C. M., Aden, D. & Koprowski, H. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1397–1400 (1975).
11. Croce, C. M., Girardi, A. J. & Koprowski, H. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3617–3620 (1973).
12. Hynes, R. O. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3170–3174 (1973).
13. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
14. Springer, T. A., Mann, D. L., DeFranco, A. L. & Strominger, J. L. *J. biol. Chem.* **252**, 4682–4693 (1977).
15. Bernoco, D., Cullen, S., Scudeller, G., Trinchieri, G. & Ceppellini, R. in *Histo-compatibility Testing* (eds Dausset, J. & Colombani, J.) 527–537 (Munksgaard, Copenhagen, 1972).
16. Hauptfeld, V., Hauptfeld, M. & Klein, J. *J. exp. Med.* **141**, 1047–1057 (1975).
17. Carnaud, C., Fadai-Ghotbi, M., Lesavre, P. & Bach, J. I. *Eur. J. Immun.* **7**, 81–85 (1977).
18. Burakoff, S. J., Dorf, M. E. & Benacerraf, B. *J. Immun.* **118**, 1795–1798 (1977).

Lymphocyte induced stimulation of human chorionic gonadotrophin production by trophoblastic cells *in vitro*

STIMULATION of target cells following an immunological interaction has been shown to occur under suitable conditions. For example, B lymphocytes undergo blast transformation when cultured in the presence of low levels of anti-immunoglobulin¹, and likewise thyroid cells can be stimulated by LATS (long acting thyroid stimulator) and LATS protector^{2,3}, both of which are immunoglobulins, resulting in increased thyroxine production. There is no evidence that target cells can be stimulated in such a manner after contact with lymphocytes. We present here evidence that increased chorionic gonadotrophin (HCG) production results when trophoblasts are cultured in the presence of maternal lymphocytes.

Trophoblastic cells were obtained from trypsinised fresh placentae and incubated for 1–3 d in Eagle's Dulbecco Modified Medium with 10% foetal calf serum in the presence or absence of lymphocytes. Lymphocytes were obtained from either normal donors, from patients with gynaecological cancer, or from pregnant women at delivery. The lymphocytes were separated on Hypaque-Ficoll and added to the trophoblast cell cultures in the ratio of 1:1 or 5:1. Production of HCG by the cultures was assessed by a radioimmunoassay technique as performed in the section of biochemistry of this hospital.

When trophoblast cells were cultured alone, there was a detectable increase in HCG production after 24 h, followed by a continuous increase over the first 3 d (Table 1). When lymphocytes from normal donors were added to trophoblast cells at a ratio of 5:1 there was slight reduction of hormone production, significant after 3 d ($P < 0.01$, Student *t* test). When lymphocytes from patients with gynaecological cancer were

Table 1 % Increase in HCG production by trophoblast cells cultured in the presence of lymphocytes (mean $\pm$ s.e.m.).

	No. of cultures	Time (Days)			
		0	1	2	3
Trophoblast only	28	100	143.4 $\pm$ 12.9	209.3 $\pm$ 9.6	248.6 $\pm$ 8.7
Trophoblast + normal lymphocytes 5:1	14	100	175.5 $\pm$ 19.4	189.6 $\pm$ 6.1	193.7 $\pm$ 10.3
Trophoblast + maternal lymphocytes 5:1	13	100	164 $\pm$ 8.9	236.5 $\pm$ 7.5	316.5 $\pm$ 16.5
Trophoblast + maternal lymphocytes 1:1	4	100	159.9 $\pm$ 12.8	191.6 $\pm$ 18.7	—
Trophoblast + cancer lymphocytes 5:1	10	100	156.6 $\pm$ 17.5	190.2 $\pm$ 20.3	185 $\pm$ 10
Trophoblast + unrelated maternal lymphocytes 5:1	2	100	—	188	—

added to the trophoblastic cells there was also significant reduction in hormone production only after 3 d ($P < 0.01$). When maternal lymphocytes taken at the time of delivery by Caesarean section were added to the trophoblastic cells at a ratio of 1:1 there was no significant difference, but when the ratio was 5:1 there was a significant increase in the amount of hormone produced after 2d ($P < 0.05$) and 3 d ($P < 0.01$). Comparison of the values obtained from cultures with added normal lymphocytes and those of cultures with added maternal lymphocytes shows a high degree of difference after 2 and 3 d ($P < 0.001$). No stimulation occurred when maternal lymphocytes from unrelated donors were used.

The reduction of HCG production by trophoblastic cells after culture with allogeneic lymphocytes is consistent with the findings of Currie⁴, Douthwaite and Urbach⁵, and Taylor and Hancock⁶, who described lysis of trophoblastic cells by allogeneic leukocytes. This 'allogeneic inhibition' occurs after 24–48 h in culture, and pre-sensitisation of the lymphocytes by target-type antigens is not required. The stimulation of HCG production by maternal lymphocytes is more difficult to explain. This phenomenon may be equivalent to stimulation of target cells by immunoglobulin, as suggested by Rose⁷ for thyroid cell stimulation by lymphoid cells. Further studies are in progress in order to establish whether this phenomenon is in any way altered in abnormal pregnancy, such as pre-eclampsia.

We thank the obstetricians and nursing staff of the Royal Women's Hospital, Melbourne, for placenta and blood samples, and Dr I. Horacek and Staff, Section of Biochemistry of this Hospital, for help with the radioimmunoassay for HCG.

W. J. DICKMAN
M. N. CAUCHI

Section of Haematology,
Royal Women's Hospital,
Melbourne, Victoria,
Australia

Received 24 October; accepted 20 November 1977.

1. Unanue, E. R., Ault, K. A., Schreiner, G. F. & Sidman, C. L. in *Proceedings of International Symposium on Membrane Receptors of Lymphocytes* (eds Seligman, M., Preud'homme, J. L. & Kourilsky, F. M.) (Paris, 1975).
2. Smith, B. R. & Hall, R. *Lancet* II, 427 (1974).
3. Adams, D. D. *et al. J. clin. Endocr.* 39, 826 (1974).
4. Currie, G. A. *J. Obstet. Gynaec. Br. Communw.* 74, 841 (1967).
5. Douthwaite, R. M. & Urbach, G. I. *Am. J. Obstet. Gynaec.* 109, 1023 (1971).
6. Taylor, P. V. & Hancock, K. W. *Immunology* 28, 973 (1975).
7. Rose, N. R. in *Basic and Clinical Immunology* (eds Fudenberg, H. H., Stiles, D. P., Caldwell, J. L. & Well, J. V.) 554 (1976).

Electrophoretic mobility as a sensitive probe of lectin-lymphocyte interaction

TREATMENT of murine T lymphocytes with concanavalin A (con A) at concentrations as low as 10^{-17} g ml⁻¹ produces significant increases in their electrophoretic mobility. The effect is blocked by α -methyl-D-mannopyranoside (α -MM). Phytohaemagglutinin (PHA) produces similar changes in T-cell mobility while treatment with lipopolysaccharide (LPS), a

B-lymphocyte mitogen, does not. We present here evidence indicating that the altered mobility of con A-treated T cells is due to the rapid release of a factor from the few cells reacting directly with the lectin.

The blast transformation of lymphocytes resulting from treatment with lectins such as con A is of considerable interest since it may serve as a model for the events which occur when lymphocytes are specifically stimulated by antigens. Lectin-induced capping, increased ³H-thymidine uptake and development of blast morphology are relatively slow phenomena. Since the earliest events associated with lymphocyte activation probably occur on the cell membrane within minutes, a technique that would allow immediate assessment of membrane alterations would be extremely useful.

Cell electrophoresis provides a probe of membrane interactions involving changes in the surface charge density, these changes being reflected by an alteration of the cell's electrophoretic mobility¹. Cell electrophoresis has previously been used as a preparative method for separating T and B lymphocyte populations^{2–6}, as a probe for lymphocyte membrane chemical composition⁷ and for studying lymphocyte interaction with a variety of chemical agents including lectins^{8–10}, polynucleotides¹¹, and antisera^{12,13}.

In view of the probable conformational changes in the lymphocyte membrane associated with transformation, it has been suspected that changes in the charge density associated with the membrane would result in an alteration in the electrophoretic mobility of the cell. We have found that extremely low concentrations of con A induce significant increases in the electrophoretic mobility of murine T cells. Moreover, on exposure to con A there is apparently a rapid release of a factor from some murine T cells which induces increased mobility in the remainder of the cells.

Electrophoretic measurements were performed using the technique of Seaman *et al.*¹ in which a suspension of cells in a controlled potential gradient is directly observed through a microscope. The results are expressed in $\mu\text{m s}^{-1}\text{V}^{-1}\text{cm} \pm 1\text{s.e.m.}$ and are based on the time of migration of a series of cells across a fixed number of divisions of an eyepiece graticule. All mobility measurements were carried out at 25 °C in 0.01 M phosphate-buffered saline, pH 7.2, using a microinjection technique to introduce cell suspensions (20–30 μl) into the observation region of the electrophoresis chamber. In most experiments, measurements were collected for 20 different cells, with a reversal of the polarity of the field between measurements. Mobility values of human and murine erythrocytes were -1.08 ± 0.02 and $1.05 \pm 0.02 \mu\text{m s}^{-1}\text{V}^{-1}\text{cm}$ which is in agreement with values reported by others^{1,14}.

Ether-anaesthetised LAF₁ mice (Jackson Laboratories) were killed and their spleens removed. Cell suspensions were prepared and washed three times in 0.01 M phosphate-buffered saline containing 5% heat-inactivated foetal calf serum (PBS-FCS), pH 7.2, by centrifugation at 490 g for 10 min at 4 °C. An enriched T-lymphocyte suspension was prepared by the method of Julius *et al.*¹⁵ using nylon wool columns (Leuko-Pak Leukocyte Filters, Fenwal Laboratories).

Con A (Miles-Yeda) was obtained as a three-times re-

crystallised powder and a stock solution ($100 \mu\text{g ml}^{-1}$) was prepared by weighing out 10 mg of con A and diluting to 100 ml with PBS. Stock con A solutions were stored at -20°C in 1 ml aliquots. Additional dilutions of con A were made by serial dilutions in PBS. α -Methyl-D-mannopyranoside (Sigma) was dissolved in PBS to yield a concentration of 0.2 M.

Phytohaemagglutinin-P (PHA) and *Salmonella typhimurium* lipopolysaccharide (LPS) (both Difco) were dissolved in distilled water. Stock solutions of PHA (0.1 ml) and LPS ($100 \mu\text{g ml}^{-1}$) were kept frozen at -20°C .

Each experiment was performed by mixing 100 μl of lectin solution or PBS and 100 μl of cell suspension in a test tube and incubating at 37°C for 5 min. Reaction mixtures containing α -methyl mannose (α -MM) were prepared by adding 100 μl of cells to a mixture of 50 μl of 0.2 M α -MM and 50 μl of lectin solution. In experiments where PHA or LPS were tested, the stock PHA and LPS solutions were diluted with PBS to a concentration of 1 : 1000 and 0.25 $\mu\text{g ml}^{-1}$ respectively.

Our initial experiments were performed using con A at final concentrations between 0.1 and $5 \mu\text{g ml}^{-1}$. A significant increase was observed (Fig. 1) in the mobility of T cells treated with con A compared with untreated T cells.

To determine whether this effect depends on con A binding to the lectin-specific carbohydrate moieties on the membrane, we repeated the experiments adding α -MM to the reaction mixture at a final concentration of 0.05 M. This sugar completely eliminated the con A-induced mobility change (Fig. 1). We excluded the possibility that the mere presence of con A in the supporting electrolyte (in contradistinction to its effect on the membrane) led to the changes in mobility. Following incubation of the T cells with con A $0.1 \mu\text{g ml}^{-1}$, we centrifuged the incubation mixture, removed the supernate and resuspended the cells in PBS. The observed mobility of -1.37 agreed well with that of the cells in the presence of the same concentration of con A and suggested that the con A effect is indeed attributable to its binding to the membrane or to induced changes in the membrane. Mitogen-induced increases in the mobility of T cells was also observed with a 1 : 1000 dilution of PHA (-1.35 ± 0.03). However, the B cell mitogen LPS, at a concentration that induces lymphocyte transformation of B cells ($0.05 \mu\text{g ml}^{-1}$) failed to induce a significant change in the mobility (-1.07 ± 0.04) of T cells.

At con A concentrations greater than $5 \mu\text{g ml}^{-1}$, a dose-dependent decrease in T-cell mobility was observed (Fig. 1). This is consistent with observations made in high concentration ranges by other investigators⁸⁻¹⁰. It can be seen (Fig. 1) that at concentrations of 25 and $50 \mu\text{g ml}^{-1}$, the T-cell mobility was actually less than that of the untreated cells. This can be interpreted as the consequence of the physical presence of large numbers of molecules of con A on the cell surface. The fact that at a con A concentration of $25 \mu\text{g ml}^{-1}$, α -MM returned the mobility to untreated levels (Fig. 1) supports this hypothesis.

Fig. 1 Influence of con A on the electrophoretic mobility of murine T lymphocytes. White bars refer to cell mobility in the absence of α -MM and black bars refer to cell mobility in the presence of α -MM. The standard error of the mean is given for the number of measurements indicated above each bar.

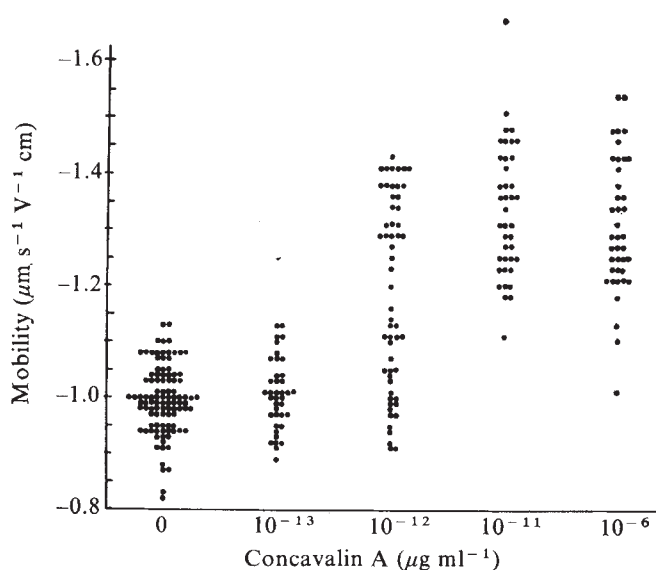
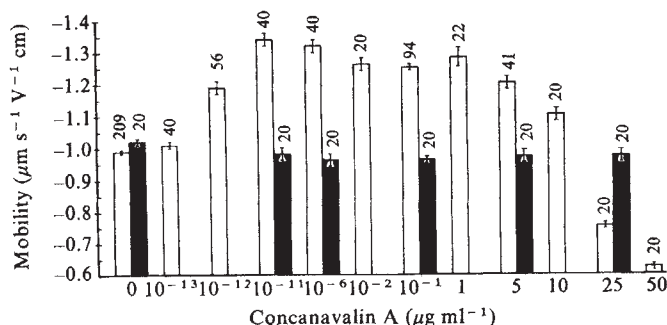


Fig. 2 Distributions of mobilities of individual cells treated with various concentrations of con A. Note that at the threshold concentration of $10^{-12} \mu\text{g ml}^{-1}$ of con A the mobility values of individual cells are distributed from control values to the maximum response values.

Furthermore, far from contradicting the results obtained at low concentrations, these data suggest that the increased mobility observed in those experiments can indeed be attributed to induced alteration of the membrane as opposed to the simple presence of a foreign molecule on it. The remaining experiments dealt with this hypothesis.

Decreasing concentrations of con A were incubated with T cells and mobility changes were clearly evident down to a concentration of $10^{-11} \mu\text{g ml}^{-1}$ (Fig. 1). At $10^{-11} \mu\text{g ml}^{-1}$ a maximal increase in electrophoretic mobility was observed with indications that the bulk of the cell population was responding, that is, on observation of the cell population in the electrophoresis apparatus there was no evidence of heterogeneity in the electrophoretic properties of the population. However, at $10^{-12} \mu\text{g ml}^{-1}$ a smaller increase in the mean mobility of the population was observed (-1.19 ± 0.02) and electrophoretic heterogeneity in the population was apparent (Fig. 2). The limited mobility distribution data collected suggested that this lower mean mobility value represented the composite of two cell populations: one consisting of cells with increased mobility and a second characteristic of unaffected cells (Fig. 2). Concentrations of con A below $10^{-12} \mu\text{g ml}^{-1}$ do not induce changes in the mobility of T cells.

The general T cell population response at the threshold con A concentration of $10^{-11} \mu\text{g ml}^{-1}$ was difficult to explain in terms of a direct interaction between lectin molecules and the surface of each lymphocyte in the population. Even if we assume the con A to exist as a monomer having a molecular weight of 27,000, at a threshold concentration of $10^{-11} \mu\text{g ml}^{-1}$, only 220 molecules of con A are necessary to initiate the observed mobility changes in 10^6 T cells. The short incubation time and the few molecules of con A needed to initiate the observed mobility changes suggested that factors distinct from con A binding led to the general T cell response. In order to determine whether con A binding of cells releases a soluble factor(s) which influences the mobility of non-con A binding cells, the following experiment was carried out.

Con A ($10^{-6} \mu\text{g ml}^{-1}$) was incubated with 10^6 T cells for 5 min at 37°C . The cell suspension was centrifuged at 1,000 g for 5 min at room temperature and 50 μl of the supernate was mixed with 50 μl of α -MM (final concentration 0.05 M) and 100 μl of untreated T cells (10^6) was added. This mixture was incubated for 5 min at 37°C and the mean mobility of the cells was -1.30 ± 0.02 . Supernate obtained from $10^{-6} \mu\text{g ml}^{-1}$ con A incubated without T cells for 5 min at 37°C , centrifuged

at 1,000g for 5 min, and added to 10^6 untreated cells in the presence of α -MM had a mean mobility of -1.00 ± 0.02 .

These observations indicate that con A treatment of T cells rapidly induces the release of a factor which can produce an electrophoretic mobility increase in the murine T cell population. Since electrophoresis is sensitive to the charge density on the cell surface, a marked alteration in the cell surface organisation of the T lymphocytes is indicated. Frelinger *et al.*¹⁶ reported that only a small portion of T cells directly respond to con A. They demonstrated that the con A responding T cells were Ia positive and release a factor that can promote the mitogenic response of Ia-negative T cells.

Studies are in progress to clarify the relationship between the factor which induces changes in T-cell mobility and the mitogenic-promoting factor described by Frelinger *et al.*¹⁶ and to establish whether similar electrophoretic responses and factor(s) occur during the interaction of immune T cells with antigens.

PHILIP BLUME

Department of Pathology,
Good Samaritan Medical Center,
Portland, Oregon 97210

ARTHUR MALLEY

Immune Disease Department,
Oregon Regional Primate Research Center,
Beaverton, Oregon 97005

ROBERT J. KNOX
GEOFFREY V. F. SEAMAN

Department of Neurology,
University of Oregon Health Sciences Center,
Portland, Oregon 97201

Received 12 September; accepted 14 November 1977.

- Seaman, G. V. F. in *The Red Blood Cell* Vol. 2 (eds Surgenor, D. Mac N.) 1135-1229 (Academic, New York, 1975).
- Zeiller, K. *Behring Inst. Mitt.* 52, 11-25 (1972).
- Stein, G. *Biomedicine* 23, 5-11 (1975).
- Nordling, S., Andersson, L. C. & Häyry, P. *Eur. J. Immun.* 2, 405-410 (1972).
- Andersson, L. C., Nordling, L. C. & Häyry, P. *Cell. Immun.* 8, 235-248 (1973).
- Thomas, C. B. & Cater, D. B. *Br. J. exp. Path.* 53, 111-124 (1972).
- Mehrishi, J. N. & Zeiller, K. *Br. med. J.* 1, 360-362 (1974).
- Wioland, M. *Cell Immun.* 12, 472-475 (1974).
- Holtz, G. & Kühn, P. *Z. Immun.-Forsch.* 144, 11-17 (1972).
- Currie, G. A. *Nature* 216, 694-695 (1967).
- Hanjan, S. N. S. & Talwar, G. P. *J. Immun.* 114, 55-58 (1975).
- Mitchell, D. M. & Cater, D. B. *Br. J. exp. Path.* 52, 152-171 (1971).
- Wiig, J. N. *Acta path. microbiol. scand. C* 84, 177-184 (1976).
- Bangham, A. D., Flemans, R., Heard, D. H. & Seaman, G. V. F. *Nature* 182, 642-644 (1958).
- Julius, M. H., Simpson, E. & Herzenberg, L. A. *Eur. J. Immun.* 3, 645-649 (1973).
- Frelinger, J. A., Neiderhuber, J. E. & Shreffler, D. C. *J. exp. Med.* 144, 1141-1146 (1976).

Amphetamine induces β -adrenergic receptor supersensitivity

THE interactions of amphetamines and catecholamines in the peripheral sympathetic nervous system and in the brain have been widely studied. Amphetamines have been reported to mimic catecholamines at their receptor sites¹⁻³, inhibit monoamine oxidase⁴, impair reuptake mechanism for the catecholamines⁵ and to directly release catecholamines into the synaptic cleft⁶. Although the consensus of the available literature indicates that *d*-amphetamine indirectly stimulates catecholamine post-synaptic receptors by increasing the release and blocking the reuptake of catecholamines^{7,8}, the question whether dopamine, noradrenaline or both, are of importance for the central actions of *d*-amphetamine is controversial^{7,8}. Recently, α -adrenergic⁹, β -adrenergic¹⁰⁻¹³ and dopaminergic^{14,15} receptors in brain tissue have been successfully identified by measuring the binding of radiolabelled ligands to specific receptor sites. Availability of such methods has permitted the examination of the effects of acute and chronic administration of psychotropic drugs on the postsynaptic catecholaminergic receptors in brain tissue^{13,16}. We report here the effects of acute and chronic

administration of *d*-amphetamine on the postsynaptic β -adrenergic receptors in rat brain.

Specific binding of ³H-dihydroalprenolol to the brain particulate fractions obtained from control rats and from rats treated with a single dose of *d*-amphetamine (10 mg per kg body weight) was measured. The binding of ³H-dihydroalprenolol to specific β -adrenergic receptors¹⁰⁻¹³, did not significantly change at 1, 12, 24, 48 and 96 h after administration of a single dose of *d*-amphetamine (results not shown). Thus, acute treatment with *d*-amphetamine did not significantly alter the sensitivity of β -adrenergic receptors in rat brain. To evaluate the effects of chronic treatment with *d*-amphetamine on rat brain β -adrenergic receptors, specific binding of ³H-dihydroalprenolol was measured 1, 12, 24, 48 and 96 h after the last dose of the drug. Although increased specific binding of ³H-dihydroalprenolol to rat brain particulate fraction could be seen 1 h after discontinuation of chronic treatment, the increase was not significant when compared to that of control (Fig. 1). Maximal activation of β -adrenergic receptor binding was observed 12 h after cessation of *d*-amphetamine injection. Specific binding of ³H-dihydroalprenolol at 12, 24, and 48 h after the last injection was significantly greater than that observed in the control preparation. According to Lewander¹⁷, rats remain in a withdrawal state for 4 to 5 d after termination of chronic treatment with amphetamine. Similarly, 96 h after cessation of chronic treatment with *d*-amphetamine, specific ³H-dihydroalprenolol binding to rat brain particulate fraction

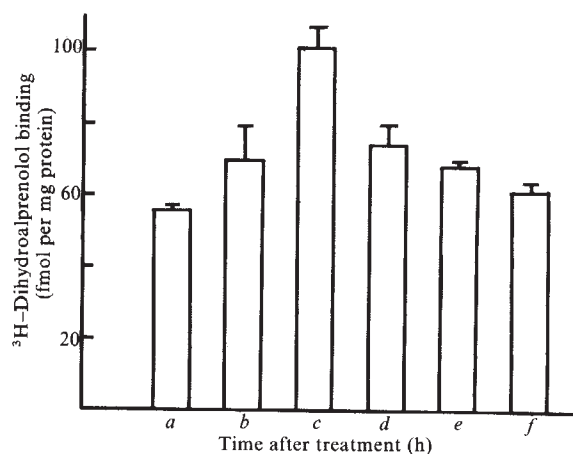


Fig. 1 Effect of chronic administration of *d*-amphetamine on specific binding of ³H-dihydroalprenolol. The drug was administered intraperitoneally at a dose of 10 mg per kg daily for a period of six weeks. The rats were killed at the indicated times after drug administration by decapitation and the brains were rapidly removed and chilled in ice-cold 0.9% NaCl. The whole brain minus cerebellum was homogenised with a Brinkman Polytron, in 20 volumes of ice-cold 50 mM Tris buffer (pH 8.0 at 25 °C), and centrifuged at 49,000g for 15 min. The pellet was homogenised in the same buffer and centrifuged as before. The pellet was finally resuspended in 50 volumes of 50 mM Tris buffer and used fresh for the binding assay. ³H-Dihydroalprenolol (32 Ci mmol⁻¹) binding was determined at 25 °C by incubating 0.97 ml of the brain particulate suspension, 2.0 nM ³H-dihydroalprenolol and sufficient distilled water to make 1 ml as the final volume of the reaction mixture. After 20 min incubation, the reaction mixtures were filtered under reduced pressure through Whatman glass fibre filters (GF/B). The filters were rinsed four times with 4 ml ice-cold Tris buffer to remove most of the unbound radioactive ligand. The filter papers were dried and placed in a counting vial containing 10 ml of scintiverse (Fisher) and counted in a Packard Tri-carb liquid scintillation spectrometer (Model 3380). Corrections were made for nonspecific accumulation of radioactivity by assaying parallel incubations which contained large excess (100 μ M) of noradrenaline. Specific binding defined as the difference between total and nonspecific radioactivity was 60-70% of the total radioactive counts. Results shown are averages of nine determinations from three rat brains at each time point, except control where six animals were used. S.e.m. values are indicated by the bars. This experiment has been replicated twice. a, Control; b, 1 h ($P > 0.05$); c, 12 h ($P < 0.001$); d, 24 h ($P < 0.01$); e, 48 h ($P < 0.001$); f, 96 h ($P > 0.05$).

was not significantly different from that obtained with the control preparations (Fig. 1). This would suggest that β -adrenergic receptor supersensitivity may be seen during the withdrawal state following chronic treatment with *d*-amphetamine. Nevertheless, more evidence is needed in support of this possibility.

Both the selective augmentations of motor behaviours in animals and the psychosis in humans associated with chronic amphetamine abuse may reflect a change in catecholaminergic receptor sensitivity. Ellinwood *et al.*¹⁸ proposed that the substantial depletion of catecholamines that accompanies long-term amphetamine treatment allows certain central catecholamine receptors to become supersensitive. This may account for the heightened psychomotor stimulation produced by subsequent injections of amphetamine. The results shown in Fig. 1 are the first biochemical demonstration of the central β -adrenergic receptor supersensitivity after chronic amphetamine intoxication.

Progressive changes in hyperactivity and stereotyped behaviour as a function of the multiple administration of *d*-amphetamine have been considered to be due to the development of reverse tolerance^{19–21}. In guinea pigs²² and in rats²³ readministration of *d*-amphetamine after cessation of daily drug treatment results in a faster onset and also an increase in the intensity of stereotyped behaviour, while administration of apomorphine only affects the faster onset of biological response. Since apomorphine is known to be an agonist of dopamine receptors, these results may suggest that supersensitivity of dopaminergic receptors is related to the decreased latency only and probably does not account for all indices of reversed tolerance. On the other hand, since there is some evidence of involvement of the noradrenergic system in amphetamine induced stereotyped behaviour²⁰, it is suggested that supersensitivity of β -adrenergic receptors may contribute, in part, to the enhancement of the intensity of stereotyped behaviour during chronic amphetamine intoxication.

Thus, chronic treatment with *d*-amphetamine causes a supersensitivity of β -adrenergic receptors in rat brain. This supersensitivity of β -adrenergic receptors may be involved, in part, in the withdrawal reaction to the chronic administration of the drug, tolerance to anorectic effects of amphetamines and enhancement of the intensity of stereotyped behaviour which may lead to psychosis in humans.

This work was supported in part by grants from the US NIH and the AHA, and a postdoctoral fellowship (to V.K.S.) from the Genesee Valley Heart Association.

SHAILESH P. BANERJEE
VIRENDRA K. SHARMA
LILY S. KUNG

Departments of Pharmacology and Toxicology,
University of Rochester, School of Medicine and Dentistry,
Rochester, New York 14642

SUBIR K. CHANDA

Department of Pharmacology,
Pennwalt Pharmaceutical Division,
Pennwalt Corporation,
Rochester, New York 14623

Received 29 September; accepted 21 November 1977.

1. Brodie, B. B. & Shore, P. A. *Ann. N. Y. Acad. Sci.* **66**, 631–642 (1957).
2. Smith, C. B. *Pharmac. exp. Ther.* **142**, 343–350 (1963).
3. Van Rossum, J. M., VanDerSchoot, J. B. & Hurkmans, J. A. T. M. *Experientia* **20** (18), 229–231 (1962).
4. Mann, P. J. G. & Quastel, J. H. *Biochem. J.* **34**, 414–431 (1940).
5. Snyder, S. H. *Archs gen. Psychiat.* **27**, 169–179 (1972).
6. Stein, L. *Fed Proc.* **23**, 836–850 (1964).
7. Groves, P. M. & Rebec, G. V. *A. Rev. Psychol.* **27**, 91–128 (1976).
8. Bunney, B. S., Walters, J. R., Kuhar, M. J., Rolti, R. H. & Aghajanian, G. K. *Psychopharmac. Commun.* **1**, 177–190 (1975).
9. U'Prichard, D. C., Greenberg, D. A. & Snyder, S. H. *Molec. Pharmac.* **13**, 454–473, 1977.
10. Alexander, R. W., Davis, J. N. & Lefkowitz, R. J. *Nature* **258**, 437–440 (1975).
11. Nahorski, S. R. *Nature* **259**, 488–489 (1975).
12. Bylund, D. B. & Snyder, S. H. *Molec. Pharmac.* **12**, 568–580 (1976).
13. Banerjee, S. P., Kung, L. S., Riggi, S. J. & Chanda, S. K. *Nature* **268**, 455–456 (1977).
14. Seeman, P., Chou-Wong, M., Tedesco, J. & Wong, K. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4376–4380 (1975).
15. Burt, D. R., Enna, S. J., Creese, I. & Snyder, S. H. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4655–4659 (1975).
16. Burt, D. R., Creese, I. & Snyder, S. H. *Science* **196**, 326–327 (1977).
17. Lewander, T. *Psychopharmacologia* **13**, 394–402 (1968).
18. Ellinwood, Jr, E. H., Sudilovsky, A. & Nelson, L. M. *Am. J. Psychiat.* **130**, 1088–1093 (1973).
19. Magos, L. *Eur. J. Pharmac.* **6**, 200–201 (1969).
20. Segal, D. J. & Mandell, A. J. *Pharmac. Biochem. Behav.* **2**, 249–255 (1974).
21. Ellinwood, E. H. & Balster, R. L. *Eur. J. Pharmac.* **28**, 35–41 (1974).
22. Klawans, H. L. & Margolin, D. I. *Archs. gen. Psychia.* **32**, 725–732 (1975).
23. Kilbey, M. M. & Ellinwood, E. H. *Life Sci.* **20**, 1063–1076 (1977).

Antibodies to thaumatin as a model of the sweet taste receptor

In developing a radioimmunoassay for thaumatin, a sweet-tasting plant protein, we have found that antibodies raised against the molecule cross-react with a wide variety of non-protein sweet substances. The results suggest that some structural feature of thaumatin which interacts with the sweet-taste receptor is also the major antigenic determinant. We suggest that antibodies raised against thaumatin may provide a novel method for measuring 'sweetness' *in vitro* and may throw light on the nature of interactions between sweet substances and taste receptors.

Thaumatin occurs in the aril of fruit of the West African plant *Thaumatococcus danielli* (Benth). Although the sweet properties were first reported in 1855¹, the active principle was not identified as a protein until recently^{2,3}. There are at least two proteins, designated as thaumatin I and II, which differ

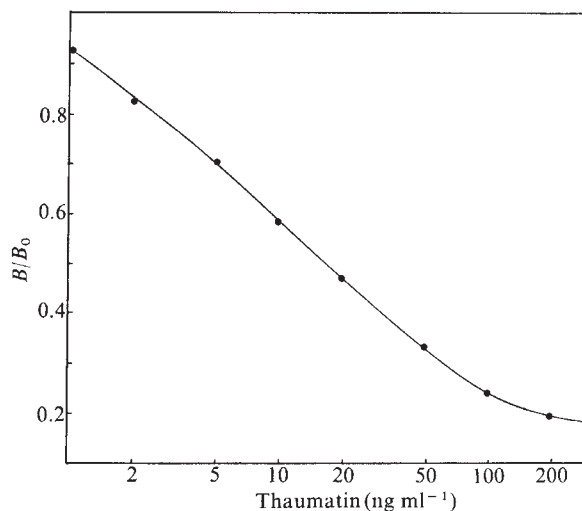


Fig. 1 Standard curve for thaumatin using the coated-tube radioimmunoassay. Thaumatin I, extracted and purified by the method of Higginbotham⁴, was used to raise antisera. Three adult male New Zealand white rabbits each received multiple intradermal injections of a total of 4.0 mg of thaumatin in Freund's complete adjuvant. After a single booster injection at 12 weeks suitable antisera were obtained. Purified thaumatin was labelled with ¹²⁵I in a 1:1 molar ratio using either chloramine-T⁵ or lactoperoxidase⁶, both techniques giving satisfactory incorporation. The labelled protein was separated from unreacted iodine by gel filtration using Sephadex G-50 (fine). A solid phase radioimmunoassay was developed with antibody-coated tubes, using a method previously described⁷ for peptide hormones. Antiserum was diluted in 0.05 M carbonate buffer (pH 9.6), 0.3 ml aliquots were placed in 1-ml plastic tubes for 2 h at 4 °C and the tubes were then aspirated, washed and allowed to dry. An antiserum dilution of 1:50,000 deposited sufficient antibody to bind approximately 50% of 3 pg of ¹²⁵I-labelled thaumatin in 200 μ l of 0.02 M citric acid-phosphate buffer (pH 5.8) containing 0.15 M NaCl and 0.1% bovine serum albumin. The assay was carried out at room temperature and the antibody-bound and free labelled thaumatin were separated by aspiration and washing, after incubation overnight. Increasing concentrations of unlabelled thaumatin displaced the binding of labelled thaumatin, as shown. B, Counts bound to antibody, B₀ = counts bound to antibody in the absence of unlabelled thaumatin. Nonspecific binding to 'uncoated' tubes was 0.03.

only slightly in composition and structure. After purification⁴ their sweetness is approximately 2×10^5 times sweeter than sucrose, when compared on a molar basis.

A solid-phase radioimmunoassay procedure was developed, and Fig. 1 illustrates a typical standard curve. The binding of labelled thaumatin to the antibody coated tubes was readily displaced by the addition of unlabelled thaumatin and the nonspecific binding of labelled thaumatin to tubes without antibody was small. The system thus provides a simple and sensitive procedure for the measurement of thaumatin in biochemical studies of this protein or for quality control where thaumatin is used as a food additive in low, palatable concentrations.

The finding that other non-protein sweet substances cross-reacted with the antisera arose by chance when the assay was used with a soft-drink formulation which contained 5.5×10^{-4} M sodium saccharin in the absence of thaumatin and where significant displacement was observed. A wide variety of sweet substances were tested for their ability to displace labelled thaumatin in the assay, including aspartame (1-aspartyl-1-phenylalanyl methyl ester⁸), calcium cyclamate, neohesperidin dihydrochalcone⁹, 4,1',6,6'-Cl₄-galactosucrose¹⁰, 1',6,6'-Cl₃-sucrose¹⁰ and monellin¹¹, the sweet protein from *Dioscorea-phyllum cumminsii* (Diels). All of these sweet substances caused significant displacement in the thaumatin assay (Table 1) but did not produce similar interference when tested at much higher concentrations (up to 20 mg ml⁻¹) in radioimmunoassays for luteinising hormone and corticotrophin.

Evidence that these substances were binding to antibodies raised against thaumatin was obtained when the sweet protein monellin, which has a different amino acid composition and structure from thaumatin, was labelled with ¹²⁵I and tested for binding with different concentrations of the antisera. A maximum of 35% binding was obtained when tubes coated with a 1:1,000 dilution of the antiserum against thaumatin were used to incubate 0.5 pg of labelled monellin. When the same amount of labelled monellin was incubated with the antiserum in solution and ethanol precipitation or a second antibody used to separate the bound from free labelled peptide, even greater binding (up to 85%) was observed with a 1:1,000 dilution of the antiserum and parallel binding curves were obtained with lower concentrations. Furthermore, when ¹²⁵I-labelled thaumatin was incubated with several of the sweet substances tested, including monellin, gel chromatography revealed no aggregation or other molecular transformation which might inhibit binding of thaumatin to the antibody and give the appearance of competitive inhibition.

The extent to which other sweet substances bind to antibodies raised against thaumatin can be expressed as the molar ratios required to produce equivalent displacement of labelled thaumatin in the assay and these values are shown in Table 1. Similarly, the sweetness of these various compounds relative to thaumatin can be derived from the reported values for their sweetness relative to a 10% solution of sucrose which is a

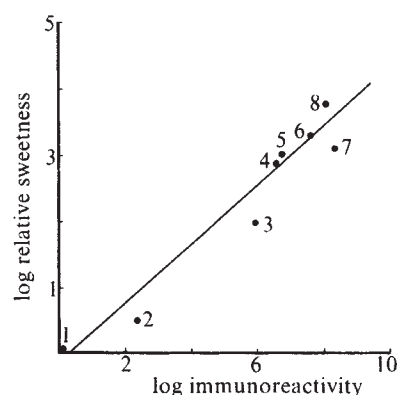


Fig. 2 Regression analysis of the relative sweetness of compounds listed in Table 1 against their immunoreactivity, both plotted on a log scale. Sweetness is expressed as the (log) molar ratio of each substance required to give sweetness equivalent to thaumatin and immunoreactivity as the molar ratio required to give equivalent displacement in the coated tube assay. Key to numbers representing each substance given in Table 1.

conventional standard for such determinations (Table 1). When the immunoreactivity and differential sweetness of these compounds are compared (Fig. 2), the sweeteners are ranked in an almost identical order, with a correlation coefficient value of 0.9606 ($P < 0.001$). It should be noted that sucrose, which is a relatively poor sweetener, will cause displacement of ¹²⁵I-thaumatin from the antibody at high concentrations (50–100 mg ml⁻¹) but at such levels nonspecific effects may influence binding to antibody.

The widely different molecular sizes and chemical structures of the sweet substances which cross-react with antibodies to thaumatin and the excellent correlation between antibody binding and sweet taste suggest that thaumatin has an outstanding structural feature which constitutes the major antigenic determinant and is also the site of interaction with the sweet taste receptor. Thus antibodies raised against this feature will cross-react with other molecules possessing physico-chemical properties which confer sweetness. Interestingly, monellin and aspartame can almost completely displace labelled thaumatin from the antibody at concentrations so low as to almost eliminate the possibility of nonspecific interactions. Although both compounds contain amino acids, the reported sequence of monellin¹² does not suggest that there is a common peptide sequence which is responsible for sweetness.

The remarkable correlation between the recognition of sweet substances by both antibodies and the sweet-taste receptor is probably coincidental but could point to some fundamental relationship in the structure of the recognition site. Binding of sweet substances to antibodies against thaumatin may provide a useful model for the study of interactions at the taste receptor for sweetness and perhaps an *in vitro* procedure for the recognition of novel sweet substances.

Table 1 Comparison of sweet substances

Sweetener	Molecular weight	Relative sweetness		Immunoreactivity
		Relative to 10% sucrose (w/w)	Relative to thaumatin (molar basis)	
(1) Thaumatin	20,000	3,000	1	1
(2) Monellin	10,700	2,000	1/3	2.4×10^2
(3) Neohesperidin dihydrochalcone	642	1,000	1/100	9.0×10^5
(4) 4,1',6,6'-Tetrachlorosucrose*	418	200	1/720	3.0×10^6
(5) L-aspartyl-L-phenylalanyl methyl ester	310	200	1/970	4.0×10^6
(6) 1',6,6'-Trichlorosucrose†	399	100	1/1,500	3.0×10^7
(7) Sodium saccharin	241	150	1/1,600	2.0×10^8
(8) Calcium cyclohexyl sulphamate (cyclamate)	200	50	1/6,000	1.0×10^8

Immunoreactivity is expressed as the molar ratios of each substance required to give equivalent displacement of ¹²⁵I-labelled thaumatin from antibody in the coated tubes assay. Relative sweetness is compared with immunoreactivity in Fig. 2.

* 'Tetrachlorosucrose' is 1',6'-dichloro-1',6', dideoxy-β-D-fructofuranosyl-4,6 dichloro-4,6 dideoxy-α-D-galactopyranoside.

† 'Trichlorosucrose' is 1',6,6'-trichloro-1',6,6'-trideoxysucrose.

We wonder also whether similar effects could occur with antisera raised against other biologically active substances; for example, could antibodies against the presumed receptor binding site for methionine-enkephalin cross-react with the non-peptide opiate drugs? Such a phenomenon would offer interesting possibilities in screening substances for pharmacological activity.

We thank Mrs Jenny Pennington and Miss Gail Webber for technical assistance, and Professor L. Hough of Queen Elizabeth College, Kensington and Dr R. Khan of Tate & Lyle Group Research and Development for providing samples of the chloro-sucroses.

C. A. M. HOUGH

Tate & Lyle Limited,
Group Research & Development,
Philip Lyle Memorial Research Laboratories,
PO Box 68,
Reading, UK

J. A. EDWARDSON

Department of Physiology,
St George's Hospital Medical School,
London SW17, UK

Received 4 October; accepted 21 November 1977.

1. Daniel, W. F. *Pharmac. J.* **14**, 158–160 (1855).
2. Van der Wel, H. & Loeve, K. *Eur. J. Biochem.* **31**, 221–225 (1972).
3. Hough, C. A. M. & Higginbotham, J. D. in *Sensory Properties of Food* (eds Birch, G. G., Brennan, J. G. & Parker, K. J.) 129–149 (Applied Science Publishers, London, 1977).
4. Higginbotham, J. D. *Belgian Patents* 84985 and 841924.
5. Hunter, W. M. & Greenwood, F. C. *Nature* **194**, 495–496 (1962).
6. Thorell, J. I. & Johansson, B. G. *Biochim. biophys. Acta* **251**, 363–369 (1971).
7. Catt, K. & Tregear, G. W. *Science* **158**, 1570–1571 (1967).
8. Mazur, R. H., Schlatter, J. M. & Goldkamp, A. H. *J. Am. Chem. Soc.* **91**, 2684–2691 (1969).
9. Horowitz, R. & Gentili, B. in *Symposium: Sweeteners* (ed. Inglett, G. E.) 182–193 (Ari, Westport, Connecticut, 1974).
10. Hough, L. & Phadnis, S. P. *Nature* **263**, 800 (1976).
11. Hough, L., Phadnis, S. P., Khan, R. A. & Jenner, M. R. *Brit. Pat. Application* No. 616/76.
12. Morris, J. A. & Cagan, R. H. *Biochim. biophys. Acta* **261**, 114–122 (1972).
13. Bohak, Z. & Li, S. H. *Biochim. biophys. Acta* **427**, 153–170 (1976).

Phospholipase A inhibition of opiate receptor binding can be reversed by albumin

OPIATE receptors in animal and human brain have been shown to be tightly associated with cell membranes^{1–3}. Our laboratory has reported the solubilisation of an etorphine–receptor complex⁴, but to date it has not been possible to obtain an active receptor in soluble form. Nevertheless, it has been possible to study many biochemical characteristics of the membrane-bound receptor. Its sensitivity to sulphhydryl reagents and a number of other protein reagents^{5,6} as well as to proteolytic enzymes^{1,7} suggests the participation of protein(s) in opiate binding. The evidence is less clear with respect to a function for phospholipids. Opiate binding activity of cell membrane preparations from rat brain has been shown to be exquisitely sensitive to phospholipase A of *Vipera russelli* venom (in the ng ml⁻¹ range)⁸, but very insensitive to phospholipase A present in the venom of *Crotalus adamanteus*¹ or that derived from pig pancreas (Lin & Simon, unpublished). Phospholipase C is inhibitory only in very high concentrations^{1,8}. The great sensitivity of opiate receptors to the Russell's viper enzyme and to ionic and non-ionic detergents^{4,7} points to a possible role for phospholipids in the binding process. We report here that inhibition of opiate binding by phospholipase A can be reversed almost completely by incubation of the enzyme-treated membranes with bovine serum albumin (BSA).

Rat brain membranes were prepared as described previously, with slight modification⁹. Male Sprague-Dawley rats were killed by decapitation. Brains were excised and cerebella were removed. Brains were homogenised in five volumes of 0.05 M Tris-buffer (pH 7.4) with a Brinkmann Polytron, at a power control setting of 5, for two 5-s intervals. The homogenate was centrifuged at 25,000g for 20 min at 4°. The pellet was resuspended in five volumes of 0.32 M sucrose and was stored

at –30° for 2 weeks without any significant loss of opiate binding activity. Frozen homogenate was thawed at room temperature and diluted with 0.05 M Tris buffer (pH 7.4) to give a final dilution of brain tissue of 1 : 60 (w/v).

V. russelli phospholipase A was placed in a boiling water bath for 15 min to destroy any contaminating proteolytic activity. The ability of the phospholipase A to inactivate opiate receptor binding was totally dependent on the presence of calcium⁷. To each reaction mixture CaCl₂ was therefore added at 2 mM concentration. Membranes were incubated with or without phospholipase A (25 ng ml⁻¹) at 37° for 10 min. The reaction was then stopped by chilling the incubate in an ice bath and the enzyme was removed by centrifugation. Phospholipase A-treated and control membranes were resuspended in the same volume of 0.05 M Tris buffer (pH 7.4) with or without 1% BSA. After incubation at 0° for 30 min, the membranes were centrifuged at 27,000g for 10 min, resuspended in the same buffer and centrifuged again. Finally, the pellets were resuspended in the same volume of Tris buffer and stereospecific binding levels of the phospholipase A-treated membranes and their controls were determined as described previously⁹, using an agonist (³H-etorphine), an antagonist (³H-naltrexone) and an endogenous opioid peptide (³H-methionine enkephalin).

The ability of phospholipase A-treated preparations to bind tritiated etorphine, naltrexone or Met-enkephalin was less than 10% that of untreated controls (Table 1). However, the same enzyme-treated membranes after incubation with BSA demonstrated a dramatic recovery of opiate binding activity. As shown in Table 1, binding of all three ligands was now 80–85% of control. Figure 1 shows that fatty acid-poor BSA elicits the same results as non-defatted albumin; other serum albumins, human, porcine, chicken and rabbit, can restore the opiate binding activity of the phospholipase A-treated membranes. Egg albumin was the only albumin tested which did not demonstrate reversal of the effects of phospholipase A. The reason for the lack of restitution of opiate binding by egg albumin is unknown. Of other serum proteins tested, α - and β -globulins demonstrated receptor reactivation of the phospholipase A-treated membranes but γ -globulins were ineffective. We have also incubated trypsin and Pronase-treated membranes with BSA, but restoration of opiate binding activity could not be demonstrated. (Lin & Simon, unpublished). This suggests that the restorative effect of BSA is specific for phospholipase A-treated membranes.

Fatty acids and lysophosphatides, the end products of phospholipolysis have also been shown to be inhibitory to the opiate binding activity of rat brain membranes, but only at relatively high (in the millimolar range) concentrations (Lin & Simon, unpublished). As BSA has been used to remove both

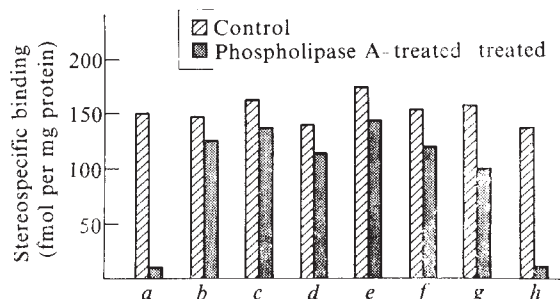


Fig. 1 Restitution of stereospecific opiate binding in phospholipase A-treated rat brain membranes by albumins. Membrane fractions were treated with phospholipase A from *Vipera russelli* venom (11 units per mg protein, Sigma). The procedures of incubation with the albumins are the same as those with BSA described in the text. The determination of stereospecific binding of ³H-etorphine as described in the legend to Table 1. a, no albumin added; b, bovine serum albumin (Pentex, Fraction V); c, bovine fatty acid-poor (Pentex, Fraction V); d, human (Pentex, Fraction V); e, porcine (Sigma, Fraction V); f, chicken (Sigma, Fraction V); g, rabbit (Sigma, Fraction V); h, egg (Sigma, crystallised).

Table 1 Effect of bovine serum albumin on phospholipase A-treated rat brain membranes

Ligand	Phospholipase A	1% BSA	Stereospecific binding (fmol per mg protein)	% of individual control
³ H-etorphine	—	—	136.7	100
	+	—	11.8	8.6
	—	+	126.3	100
	+	+	104.6	82.8
³ H-naltrexone	—	—	139.2	100
	+	—	7.9	5.7
	—	+	134.2	100
	+	+	116.5	86.8
³ H-enkephalin	—	—	61.2	100
	+	—	3.9	6.4
	—	+	78.9	100
	+	+	62.0	78.6

Membrane fractions derived from rat brain were treated with phospholipase A from *Vipera russelli* (11 units per mg protein, Sigma) and incubated with or without BSA (Pentex, Fraction V) as described in the text. Opiate receptor stereospecific binding was assayed by incubation of the phospholipase A-treated and control membranes with 1.1 nM ³H-etorphine (16.5 Ci mmol⁻¹) or 1.8 nM ³H-naltrexone (10.6 Ci mmol⁻¹) with or without 1 μM unlabelled levallorphan at 37° for 15 min. Samples were placed in an ice bath for 10 min before filtration through GF/B filters. ³H-Met-enkephalin (15.9 Ci mmol⁻¹) was used at a concentration of 4.6 nM and incubation was carried out at 4° for 1.5 h with or without 20 μM unlabelled enkephalin.

fatty acids and lysophosphatides from solutions or from membranes¹⁰, the recovery of binding activity following albumin treatment might result from the removal of the end products of phospholipase A hydrolysis. This hypothesis received some support from the observation that the addition of BSA simultaneously with phospholipase A completely prevented the inhibition of opiate receptor activity. However, the inhibition of opiate binding by added lysolecithin or linoleic acid was not restored by treatment with BSA (unpublished results). Nevertheless, an inhibitory effect elicited by the products of phospholipase A hydrolysis cannot be ruled out, because the endogenous fatty acids and lysophosphatides produced within the membrane may have effects on the opiate receptors at much lower concentrations than the exogenous materials, which were added as sonicated suspensions in buffer.

As can be seen in Table 1, the inhibitory effect of phospholipase A on the binding activity of these membranes for tritiated etorphine, naltrexone and Met-enkephalin was similar and their binding was substantially restored by incubation with BSA. These results provide evidence that the binding of agonist, antagonist and endogenous morphine-like material involves the same or similar binding sites.

The remarkable sensitivity of opiate receptors to the phospholipase A of *Vipera russelli* and the reversibility of this inhibition by albumin may provide a useful tool with which to define the role of phospholipids in the stereospecific binding of opiates and endorphins. Moreover, methods for the solubilisation of active opiate receptors may also depend ultimately on the interrelationship of phospholipids and receptor proteins.

We thank Dr P. Elsbach for helpful discussions out of which came the idea to try BSA. This work was supported by grant DA00017 from the National Institute on Drug Abuse. We acknowledge a financial gift from Hoffmann-La Roche, Inc. Nutley, N.J.

HUNG-KUANG LIN
ERIC J. SIMON

Department of Medicine,
New York University Medical Center,
New York 10016

Received 22 September; accepted 5 December 1977.

- Simon, E. J., Hiller, J. M. & Edelman, I. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1947–1949 (1973).
- Pert, C. B. & Snyder, S. H. *Science* **179**, 1011–1014 (1973).
- Terenius, L. *Acta pharmac. toxicol.* **32**, 317–320 (1973).
- Simon, E. J., Hiller, J. M. & Edelman, I. *Science* **190**, 389–390 (1975).
- Simon, E. J. & Groth, J. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2404–2407 (1975).
- Pasternak, G. W., Wilson, H. A. & Snyder, S. H. *Molec. Pharmac.* **11**, 340–351 (1975).
- Pasternak, G. W. & Snyder, S. H. *Molec. Pharmac.* **10**, 183–193 (1974).
- Pasternak, G. W. & Snyder, S. H. *Molec. Pharmac.* **11**, 478–484 (1975).
- Simon, E. J., Hiller, J. M., Groth, J. & Edelman, I. *J. Pharmac. exp. Ther.* **192**, 531–537 (1975).
- Azhar, S., Hajra, A. K. & Menon, K. M. J. *J. biol. Chem.* **251**, 7405–7412 (1976).

Calcium ion-dependent vesicle fusion in the conversion of proalbumin to albumin

PROALBUMIN is the immediate biosynthetic precursor of albumin¹, and has been shown (in the rat) to contain a N-terminal extension (Arg-Gly-Val-Phe-Arg-Arg)^{2,3} which can be cleaved by trypsin to yield albumin. We have recently found an enzyme in rat liver, tentatively identified as cathepsin B (ref. 4), which converts proalbumin to albumin, and which is inhibited 70% by leupeptin (3 μM) and by tosyl-lysyl-chloro-methyl ketone (TLCK, 10 μM). Purified bovine cathepsin B (provided by Dr A. J. Barrett) has also been found to carry out the conversion of proalbumin. Zühlke *et al.*⁵ have discussed the possibility that cathepsin B might catalyse the conversion of proinsulin; and since double basic residues occur at the cleavage points of a number of proproteins, a common enzymatic mechanism might be involved in their conversion for secretion. Our earlier work⁶ suggested that conversion of proalbumin was a very late event in the secretory pathway, which agrees with the suggestion of Ikehara *et al.*⁷ that conversion occurs in Golgi vesicles *in vivo*. Here, we present evidence that conversion takes place in isolated Golgi vesicles *in vitro*, that the enzyme concerned is probably cathepsin B, and that the conversion of proalbumin to albumin involves a Ca²⁺-requiring intracellular vesicle fusion occurring at or beyond a transit step through the Golgi apparatus.

Rats were killed 15 min after intravenous injection with ³H-valine (that is, at a time when labelled albumin would just be emerging from the liver)⁸. Liver Golgi vesicle preparations were made by a modification of the method of Merritt and Morre⁹. The vesicles were found to contain a mixture of radioactive proalbumin and albumin, the proportions of which varied from preparation to preparation and which averaged

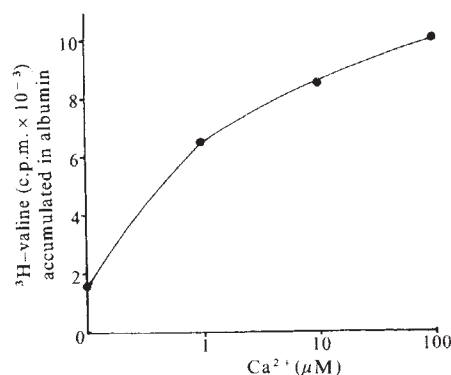


Fig. 1 Calcium requirement for conversion of proalbumin to albumin in Golgi vesicles. Golgi vesicles were separated from the livers of two rats⁹, one of which was given 300 μCi ³H-valine 15 min before death. They were finally suspended in 70 mM sucrose containing 1 mM EGTA. Ca²⁺ was added to give the free Ca²⁺ concentration indicated¹¹. The samples were incubated at pH 6.0 (50 mM 2-[N-morpholino] ethanesulphonic acid) with dithiothreitol 1 mM. Final volume 1 ml. Incubation was for 60 min at 30 °C. The reactions were stopped by addition of 200 μl M Tris-HCl, pH 7.8, 50 μl 10% (w/v) sodium deoxycholate, 750 μg unlabelled rat albumin and 5 ml 150 mM NaCl. Albumin and proalbumin were precipitated with anti-rat albumin and separated using ion-exchange chromatography¹⁷ after release from the precipitate².

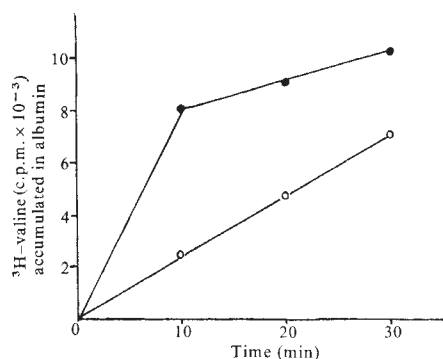


Fig. 2 Effect of a chemical fusogen on conversion of proalbumin in Golgi vesicles. The preparation of Golgi vesicles was as described in Fig. 1, except that just before the final sedimentation, half the vesicles were treated with an equal volume of 80% (w/v) PEG for 1 min at 20 °C. The remainder were treated in the same way, with an equal volume of 70 mM sucrose. They were then chilled, diluted with 70 mM sucrose and sedimented (100,000g, 20 min). They were suspended and incubated as described in Fig. 1. Ca^{2+} was added to give a concentration of 100 μM . The experiment shown is representative of four similar experiments.

about 70% in the form of proalbumin. The yield was approximately 2–3 mg protein per liver. Our Golgi vesicle preparation appeared by electron microscopy to be predominantly small vesicles, with a few cisternal stacks. Marker enzymes showed contamination by endoplasmic reticulum (5%, glucose-6-phosphatase); plasma membranes (6%, 5' nucleotidase) and lysosomes (8%, cathepsin D). No mitochondrial markers were found. The specific activity of UDP-galactose-4-epimerase was increased 50-fold over that of the homogenate. In some experiments, the Golgi vesicle fraction (GF_1) of Ehrenreich *et al.*¹⁰ was also tested. The results were identical in the two preparations.

For the experiments, the vesicles were removed from the gradient, sedimented and suspended in 70 mM sucrose containing 1 mM EGTA. The albumin and proalbumin content was determined by disruption of the vesicles with 0.5% (w/v) deoxycholate and immunoprecipitation using anti-albumin. The proteins were separated by chromatography on DEAE cellulose^{2,17}.

A large and reproducible conversion was observed when the vesicles were incubated at pH 6.0 in the presence of Ca^{2+} , the concentrations of which were maintained as described by Portzehl *et al.*¹¹ (Fig. 1). Mg^{2+} did not replace Ca^{2+} . Conversion was inhibited 50% by TLCK (10 μM) and 75% by leupeptin (10 μM added partly free and partly entrapped in multilamellar liposomes). Conversion was completely prevented if the vesicles were pretreated with 1 mM DTT and held at pH 7.2 for 30 min at 30 °C before the incubation at pH 6.0. These experiments suggest that cathepsin B or a similar enzyme was responsible for conversion. There was no loss of radioactive protein during conversion and chromatography revealed the presence only of proalbumin and albumin.

It occurred to us that the substrate (proalbumin) and the converting enzyme might be present in separate populations of vesicles which fuse in a controlled manner (see Novikoff's review¹²), an idea which was reinforced by the report that Golgi vesicles can be induced to fuse in the presence of low concentrations of Ca^{2+} (ref. 13). We therefore undertook experiments with polyethylene glycol (PEG) which has been shown to induce fusion in cells and protoplasts^{14–16}. Golgi vesicles were incubated with 40% (w/v) PEG (molecular weight 6,000) for 60 s at 20 °C, and the fusogen removed by centrifugation. There was a marked stimulation in the conversion of proalbumin to albumin in a subsequent incubation at 30 °C, as shown in Fig. 2.

If the PEG is left in contact with the vesicles, there is complete inhibition of conversion. No Ca^{2+} is required during the pretreatment with PEG, but the subsequent conversion has the

same Ca^{2+} requirement as in the absence of PEG treatment (see Fig. 1).

No conversion was seen in smooth endoplasmic reticulum fractions. A crude Golgi cisternae fraction gave only 30% of the conversion seen in Golgi vesicles separated from the same liver. In the experiments, conversion was shown to be intravesicular. Typically, after incubating 60 min in the absence of Ca^{2+} the albumin contained 6,000 c.p.m. ^3H , whereas with Ca^{2+} 100 μM the figure was 11,000. In the supernatant of both albumin was found containing 2,000 c.p.m. Zero-time controls were identical to samples incubated without Ca^{2+} .

These experiments are compatible with the notion that cathepsin B (or a similar enzyme) converts proalbumin to albumin. It is also reasonable to believe that enzyme and substrate are in separate populations of vesicles which fuse in the presence of Ca^{2+} at or near intracellular concentration.

The results suggest that the fused vesicles form a secretory vesicle, which in turn is able to recognise and fuse with the plasma membrane to allow exocytosis. This would account for the fact that proalbumin is not found in extracellular fluids. The results might also explain the action of colchicine which inhibits the conversion of proalbumin to albumin¹⁷ and of parathyroid hormone (pro-PTH) to PTH¹⁸ as well as blocking the secretion of albumin and PTH, and may therefore have a wide interest.

This work was supported by grants from the Cancer Research Campaign, the Vandervell Foundation and the Smith, Kline and French Foundation. We thank Drs R. T. Dean, A. J. Barrett and B. D. Gomperts for assistance.

J. D. JUDAH
P. S. QUINN

Department of Experimental Pathology,
University College Hospital Medical School,
University Street, London WC1, UK.

Received 3 August; accepted 9 December 1977.

- Judah, J. D., Gamble, M. & Steadman, J. H. *Biochem. J.* **134**, 1083–1091 (1973).
- Quinn, P. S., Gamble, M. & Judah, J. D. *Biochem. J.* **146**, 389–393 (1975).
- Russell, J. H. & Geller, D. M. *J. Biol. Chem.* **250**, 3409–3413 (1975).
- Dean, R. T. & Barrett, A. J. in *Essays in Biochemistry* **12** (eds Aldridge, W. N. & Campbell, P. N.) 1–40 (Academic, London, 1976).
- Zühlke, H., Steiner, D. F., Lernmark, A. & Lipsey, C. in *Polypeptide Hormones: Molecular and Cellular Aspects* Ciba Foundation Symposium, **41**, 183–191 (1976).
- Geller, D. M., Judah, J. D. & Nicholls, M. R. *Biochem. J.* **127**, 865–874 (1972).
- Ikehara, Y., Oda, K. & Kato, K. *Biochem. biophys. Res. Commun.* **72**, 319–326 (1976).
- Judah, J. D. & Nicholls, M. R. *Biochem. J.* **123**, 649–655 (1971).
- Merritt, W. D. & Morré, D. J. *Biochim. biophys. Acta* **304**, 397–407 (1973).
- Ehrenreich, J. H., Bergeron, J. J. M., Siekevitz, P. & Palade, G. E. *J. Cell Biol.* **59**, 45–72 (1973).
- Portzehl, H., Caldwell, P. C. & Ruegg, J. C. *Biochim. biophys. Acta* **79**, 581–591 (1964).
- Novikoff, A. B. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2781–2787 (1976).
- Gratzl, M. & Dahl, G. *FEBS Lett.* **62**, 142–144 (1976).
- Lucy, J. *Trends Biochem. Sci.* **2**, 17–20 (1977).
- Fodor, K. & Alföldi, L. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2147–2150 (1976).
- Schaeffer, P., Cami, B. & Hotchkiss, R. D. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2151–2155 (1976).
- Dorling, P. R., Quinn, P. S. & Judah, J. D. *Biochem. J.* **152**, 341–348 (1975).
- Kemper, B., Habener, J. F., Rich, A. & Potts, J. T. *Endocrinology* **96**, 903–912 (1975).

An unusual interaction between the target of nalidixic acid and novobiocin

BACTERIAL mutation resulting in drug resistance often has pleiotropic effects. Amongst these are alterations in susceptibilities to other drugs. Perhaps the best understood and studied of these interactions are antibiotic resistance mutations that involve genes of ribosomal proteins. The pleiotropic effects induced by ribosomal mutations are usually interpreted by interaction of different ribosomal proteins in the ribosomal organelle. For example, it has been observed that certain ribosomal mutations to neomycin-kanamycin (*nek*) resistance in *Escherichia coli* mask the phenotype of spectinomycin (*spc*) resistance¹. Similarly, there is also interaction between ribosomal ambiguity (*ram*) and the streptomycin (*str*) genes². More recently, it has been observed that there is an interaction of ribosomal (*str*) and RNA polymerase subunit (*rif*) mutations³. I describe here an interaction between mutations involving resistance against two quite different drugs, nalidixic acid and

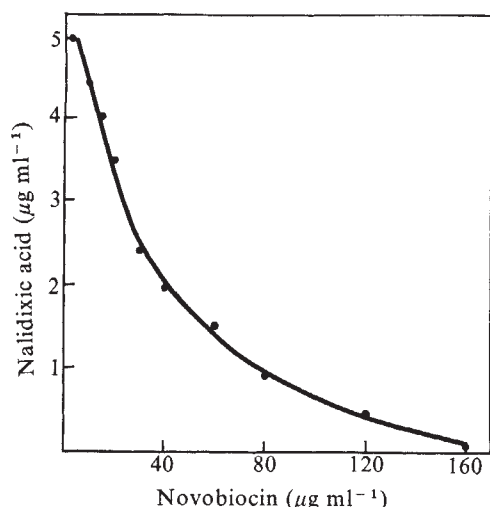


Fig. 1 Effect of drug combination on the growth of *E. coli* KL161. A series of concentrations of novobiocin in the presence of various concentrations of nalidixic acid were prepared in L broth for the drug sensitivity test. The MIC of each drug combination was determined as described in the legend to Table 1. An isobologram was constructed from these values. Each point on the isobologram represents the two drug concentrations that showed growth inhibition when used in combination.

novobiocin, both of which affect DNA replication. This genetic interaction may reflect existence of a protein complex in the DNA replication machinery which has already been postulated on other grounds (see ref. 4).

Nalidixic acid is an extensively used antibacterial drug whose precise mode of action is poorly understood. The primary action of nalidixic acid is inhibition of DNA synthesis⁵. The structurally unrelated drug, novobiocin, has been shown to inhibit DNA synthesis in a manner similar to that of nalidixic acid⁶. The most likely target of novobiocin seems to be DNA gyrase⁷. Despite a similar action of the two drugs, separate targets seem to be involved, as indicated by the distinct map location of nalidixic acid resistance⁸ and novobiocin resistance⁹ in *E. coli*. The fact that both nalidixic acid and novobiocin specifically block DNA replication suggests a possible *in vivo* interaction between the targets of these two drugs. This possibility was examined with a pair of *E. coli* strains KL161 and KL166. The two strains are isogenic except that KL166 is a *nalA* derivative of KL161¹⁰. First, drug sensitivity tests were carried out, and the results of these tests are shown in Table 1. Note that the *nalA* mutation significantly increased novobiocin sensitivity of strain KL166. This uncommon phenomenon is called collateral sensitivity¹¹. The cause of collateral sensitivity is not clear but drug-target interaction is a possibility. For example, the two drug targets may be structurally or functionally dependent on each other. A mutation in one drug target could thus impair the proper function of the other drug target. This point was further pursued with combination drug testing. It is known that drugs acting on different steps of a biochemical pathway frequently show synergism¹². The results presented in Fig. 1 clearly demonstrate a synergistic effect between the two drugs. Furthermore, in rare cases, drug targets may show such

Table 1 Sensitivity of *E. coli* KL161 and KL166 to nalidixic acid and novobiocin

Strain	MIC ($\mu\text{g ml}^{-1}$)	
	Nalidixic acid	Novobiocin
KL161	5	165
KL166	500	70

Drug sensitivity tested by broth dilution in L broth¹³. KL161 and KL166 were grown in L broth at 37 °C to 10 cells per ml. Approximately 10^3 cells from each strain were distributed into a series of tubes containing various concentrations of nalidixic acid or novobiocin. The lowest drug concentration required to arrest cell growth was designated as minimal inhibitory concentration (MIC) of that drug.

Table 2 Frequency of spontaneous mutation to drug resistance *E. coli* KL161 and KL166

Strain	Frequency of mutation to drug resistance			
	Ampicillin	Chloramphenicol	Streptomycin	Novobiocin
KL161	2.3×10^{-9}	1.8×10^{-4}	3.2×10^{-10}	4030×10^{-5}
KL166	6.5×10^{-9}	2.1×10^{-4}	1.4×10^{-10}	7.9×10^{-5}

Mutation frequency is expressed as the ratio of drug resistant colonies to the number of total viable cells. L broth and L agar were used for cell growth and for plating viable cells. L agar containing antibiotics was used for scoring spontaneously occurred drug resistant mutants. The following drug concentrations were used: ampicillin, 10 $\mu\text{g ml}^{-1}$; chloramphenicol, 5 $\mu\text{g ml}^{-1}$; streptomycin, 100 $\mu\text{g ml}^{-1}$; novobiocin, 200 $\mu\text{g ml}^{-1}$ (all drugs from Sigma).

a physical interdependence that an alteration of mutation frequency may result¹. For example, a resistance mutation to one drug could alter the frequency of resistance mutation to a second drug. This was found to be the case. The mutation frequency of the two strains was determined with several antibiotics. The results in Table 2 show that KL161 and KL166 differed in their rate of spontaneous mutation to novobiocin resistance by about 500-fold. This is a specific effect, since the frequency of mutations to ampicillin, chloramphenicol and streptomycin resistance were not significantly altered by the *nalA* mutation (Table 2). Results essentially similar to that of novobiocin were also obtained with coumermycin A₁ (data not shown). This similarity is expected since coumermycin A₁ has been shown to share a common target with novobiocin⁷. It should be noted, that the *nalA* phenotype is not masked by novobiocin resistance. Seventy-four doubly resistant mutants were selected by plating KL166 on novobiocin plates. All mutants retained their resistance to nalidixic acid when replicated on nalidixic acid plates.

Novobiocin and coumermycin have been shown to inhibit DNA gyrase activity and DNA replication in *E. coli*⁷. The precise mechanism of action of nalidixic acid is not known. It is possible that the target of nalidixic acid and that of novobiocin act independently and sequentially in DNA replication. Alternatively, the two drug targets may act interdependently in the replication complex. Taken together with published results on the actions of nalidixic acid and novobiocin, the data presented here strongly suggest that the target of nalidixic acid and novobiocin interact *in vivo*. For example, *nalA* mutation may impose a severe conformational strain on DNA gyrase. Thus, only a limited number of mutations in DNA gyrase are viable. This finding is not entirely surprising since a multicomponent complex is probably involved in DNA replication⁴. A mutation in one component could conceivably affect the function of adjacent components.

I thank J. F. Speyer, P. Sawyer and J. D. Karam for helpful suggestions. This work was supported by the General Medical and Faculty Research Appropriation 1977-78.

Note added in proof: The close relationship between the targets of nalidixic acid and novobiocin is also substantiated by two recent reports^{14,15}.

LEE CHAO

Biochemistry Department, Medical University of South Carolina, Charleston, South Carolina 29403

Received 25 October; accepted 23 November 1977.

1. Apirion, D. & Schlessinger, D. *Proc. natn. Acad. Sci. U.S.A.* **63**, 794-799 (1969).
2. Gorini, L. in *RNA Polymerase* (eds Chamberlain, M. J. & Losick, R.) 791 (Cold Spring Harbor Laboratory, New York, 1976).
3. Chakrabarti, S. L. & Gorini, L. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1157-1161 (1977).
4. Schekman, R., Weiner, A. & Kornberg, A. *Science* **186**, 987-993 (1974).
5. Cossarelli, N. R. *A. Rev. Biochem.* **46**, 641-668 (1977).
6. Staudenbauer, W. L. *J. molec. Biol.* **96**, 201-205 (1975).
7. Cellert, M., O'Dea, M. H., Itoh, T. & Tomizawa, J. *Proc. natn. Acad. Sci. U.S.A.* **73**, 4474-4478 (1976).
8. Hane, M. W. & Wood, T. H. *J. Bact.* **99**, 238-241 (1969).
9. Ryan, M. J. *Biochemistry* **15**, 3769-3777 (1976).
10. Bachmann, B. J. *Bact. Rev.* **36**, 525-557 (1972).
11. Szybalski, W. & Bryson, V. J. *Bact.* **64**, 489-499 (1952).
12. Jawetz, E. *A. Rev. Pharmac.* **8**, 151-170 (1968).
13. Bertani, G. *J. Bact.* **62**, 293-300 (1951).
14. Sugino, A., Peebles, C. L., Kreuzer, K. N. & Cozzarelli, N. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4767-4771 (1977).
15. Gellert, M., Mizuuchi, K., O'Dea, M. H., Itoh, T. & Tomizawa, J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4772-4776 (1977).

reviews

Anthropology and aerial warfare

J. Z. Young

From Apes to Warlords. The Autobiography (1904–1946) of Solly Zuckerman. By Solly Zuckerman. Pp 426. (Hamish Hamilton: London; Harper and Row: New York, 1978.) £7.95; \$17.95.

SCIENTISTS to-day probably associate Solly Zuckerman with Whitehall and the London Zoo. He has indeed served science well in both these places, but in this first volume of his autobiography he tells us about two quite different and earlier Sollys. In the first of these earlier incarnations, he laid the foundations of the knowledge that led to the Pill. In the second, he played a part in devising the strategy of the Royal Air Force. Not being either a gynaecologist or historian, I cannot judge how far these memoirs add new facts to what was already known. But it makes compulsive reading to follow how a young South African made his way into the main streams of anthropology, endocrinology and then, so surprisingly, of aerial warfare.

The significance of the work that he did on monkeys is here almost obscured by the diary of meetings with so many different sorts of people. It is difficult now to believe that before Zuckerman's observation of the swollen behind of the female baboon there was no firm knowledge of the nature of the menstrual cycle. That discovery could indeed be regarded as one of the landmarks of the history of human self-knowledge. Of course, it was not made by one man alone; but Solly's description of the baboons was certainly an important factor, although he modestly refers to it, almost in passing, as "one of several clues".

Out of the wide circle of acquaintances that he made in the 1930s came his life with the warlords. It all began from primitive attempts to find out how to protect people against air raids. He and Desmond Bernal organised themselves into an "Oxford Extra-Mural Unit of the Ministry of Home Security". He gives here literally a blow-by-blow account of how they learned about the effects of bombs on people. This section should be as fascinating for scientists and scientific historians as for soldiers and the general public. Beginning with a few experiments with rabbits and guinea pigs, it ended by contributing essential data for the formulation of the strategy



Zuckerman at Oxford in 1935



At the end of the War, 1945

At Tobruk in 1943

of vast armies and air forces. The method used at all times was a critical testing of hypotheses. The first one to be disproved was the official Medical Research Council view that blast kills by entering the mouth. Disproved by watching a bird fly away as a great gun fired and then showing that the lungs of a rabbit enclosed up to its neck in a box was not damaged by blast. The same combination of observation and experiment was then applied in the study of the results of bombing, first of one room in a lone raid on Banbury and then of hundreds of incidents at Hull and Birmingham. The results of this survey were the basis of one of the disputes about which much has been written between Tizard and Cherwell. Zuckerman claims that it has never been pointed out that the conclusions Cherwell drew from a preliminary version of their report were exactly the opposite of those reached by the authors of the survey. Unlike Cherwell, Bernal and Zuckerman did not believe that area bombing was effective. As Solly movingly concludes, the final proof came in the bombing of Vietnam where "seven million tons brought no victory—only death and destruction."

It is a fascinating story how this knowledge that they obtained came to

make these two scientists influential in the highest quarters. We have excellent glimpses of Mountbatten and Tedder, of Cherwell and Tizard, of Eisenhower and Churchill, and many more, told, it should be said with a simple directness of observation. Of course, he enjoyed meeting all these great men, but on the whole he kept his head. He has a sharp pen and his views are frank and will not always be welcome. He comments about some of the scientific advisors to Harris at Bomber Command that "not all the scientists who had been drawn into Service posts were as questioning and independent in their judgements as they could have been. On occasion, they were constrained by assumptions which uncannily fitted their master's preconceived ideas".

Among scientists, he has never been partial to anthropologists nor they to him. Nor is he much impressed by philosophers. His comment on Freddie Ayer is that he wishes that he had shown "the interest and energy necessary to become familiar with even a small part of science"; and he doubts "whether the growth of the natural sciences has been much influenced over the past 40 or 50 years by the deliberations of academic philosophers."

On the other hand, he sometimes

found it hard to believe in the badness of those about him. In the course of the debate over bombing policy, Zuckerman attended one of the mid-night meetings of the Defence Committee. Churchill read out a minute of Cherwell's describing the suggested bombing plan as "the brain child of a biologist who happened to be passing through the Mediterranean". But in spite of this and other evidence of Cherwell's ill-will and bad judgement Solly could never quite bring himself to recognise what an odious man he was, though he notes that Isaiah Berlin

"used to feel uncomfortable in the presence of the Professor. I never did understand why."

So we are introduced to many of the academics, politicians and soldiers who ruled the world in wartime. It is an invigorating story and shows how necessary it still is to persuade people of the value of applying the scientific method to human affairs. We shall wait eagerly for the post-War sequel. □

J. Z. Young is Emeritus Professor of Anatomy at University College, London, UK.

Current neuropeptide perspectives

Peptides in Neurobiology. Edited by Harold Gainer. Pp. 464. (Plenum: New York and London, 1977.) \$41.40.

THE principal aim of this book, says the Editor in his preface, is to provide a baseline of information and perspectives in the rapidly moving field of neuropeptides. In this aim I believe it succeeds, but it is never easy reading and it provides, in some of its chapters, a plethora of undigested and indigestible information, too much perhaps for even the most dedicated neurobiologist.

There are 14 chapters, by 23 contributors; and the neuropeptides are described, dissected and delimited in four principal, but discontinuous, subdivisions. In order of appearance these are Biochemical Methodology, Localisation and Analysis, Metabolism and Physiology, Pharmacology and Behaviour.

The opening chapter, by Berta Scharrer, is entitled "Historical Introduction". A great deal more than that, it describes, albeit in less than seven pages, the basic concepts of the modern discipline of neuroendocrinology of which the author, together with her late husband, are the undisputed founders. The ancestral neurone "equally endowed for the dispatch of long distance and strictly localised signals" yields to its successor, the neurosecretory neurone, capable of all degrees of variation between neuronal and glandular function. Acknowledgement of the inherent capacity of neuronal elements for hormone production is qualified by the current view that neurosecretory neurones manufacture chemical mediators to a degree greatly surpassing that of the more "conventional" neurones. Is this really true, or has this function in higher neurones

been overlooked because it has not yet been looked for?

The next four chapters (Stein on fluorecamine as a peptide reagent; Straus and Yalow on radioimmunoassay; Sternberger on immunocytochemistry; Leeman, Mroz and Carraway on the isolation of substance P and neurotensin) comprise the main section on "Methodology". Individually excellent, they are not equally germane to the general argument of the work. Usefully set down here, however, they will not need to be repeated when the neuropeptide story is extended by further instalments. Despite editorial denials, there is some degree of overlap in this section.

Brownstein's chapter (6) which follows, begins with some historical aspects of the localisation of neuropeptides and continues with a presentation of the results of microdissection and microassays, particularly as applied to the hypothalamus. Accompanying tables give the quantified distribution of LHRH, TRH, somatostatin, vasopressin and oxytocin.

Next, Reichelt and Edminson append a long list of oligopeptides found in the CNS, discussing their role and ending with the working hypothesis that they represent "the final common pathway of multisignal integration". This exciting but as yet unverified notion gives way to a chapter by Harold Gainer and his associates on the biosynthesis of neurohypophyseal peptides and neurophysin. The work reported concerns the hypothalamo-hypophyseal system of the rat, concentrating on the time course of events by which precursors and peptides are first synthesised and then transported from soma to axon and axon terminal.

A mainly methodological chapter by Marks deals with the conversion of prohormones by specific and non-specific peptidases, and useful lists of the latter are included. This very important field is both lucidly and comprehensively dealt with by the author.

A virtually complete survey of peptides in invertebrate nervous systems

is given by Nora Frontali (with Harold Gainer). The authors conclude that the relative prominence of neurosecretion in invertebrates reflects a paucity of endocrine systems and the lowly phylogenetic position of the neurone. The physiological role of neuropeptides is discussed by Jeffery Barker, and his chapter provides definitions of neurohormone and neurotransmitter, as well as listing the distinctions between them. This is a subject where complete agreement has yet to be reached, but Barker concludes that neurohormonal peptides can cause changes in neuronal membranes which are quite different from those produced by putative neurotransmitters.

The concluding three chapters are written by Cooke (on electrical activity and control of peptide release), Klee (on endogenous opiate peptides), and De Wied and Gispen (on behavioural effects). The first of these, devoted to work on the crustacean X-organ sinus gland neurosecretory system, provides evidence that all peptidergic neurones have full electrical activity. Klee, giving an up-to-date but concise account of the present status of the enkephalins and endorphins, is nevertheless moved to supply an even more up-to-date addendum. This emphasises, if emphasis is needed, the extraordinary rapid pace of enquiry in this branch of neuropeptidology. De Wied and Gispen end the book with a chapter describing their own well-documented studies in rats on the effects (shuttle box avoidance response, grooming, and so on) of the intracerebral exhibition of vasopressin and oxytocin analogues, ACTH and MSH peptides, and fragments thereof. The relationships described are complex, and not to be comprehended in a single reading.

With few exceptions, as has been stated, the individual chapters composing this book are not easy to read and there is an unavoidable lack of continuity between one chapter and another. A great merit of the book is the speed with which it has been published, a result bought at the expense of a superfluity of misprints, including an amusing inversion (p81) where the famous Italian apparatus is ascribed to Gogli.

Despite these minor criticisms, the book represents a valuable exposition of the present status of peptides in neurobiology, and it will do until it is overtaken by events and by its successor.

A. G. E. Pearse

A. G. E. Pearse is Professor of Histochemistry at the Royal Postgraduate Medical School, Hammersmith Hospital, London, UK.

Human and other chromosomes

Molecular Structure of Human Chromosomes: Chromosomes in Biology and Medicine. Edited by J. Y. Yunis. Pp. 336. (Academic: New York, San Francisco and London; 1977.) £17.40; \$24.50.

THIS volume emphasises in respect of human chromosomes the recent advances in technology for the study of the eukaryote nucleus. A collection of nine chapters by different authors, eminent in different specialities of the subject, gives a comprehensive cover, even if with some overlapping.

The first three chapters, by Yunis *et al.*, Macaya *et al.* and Steffensen, respectively, deal with the study of human and other chromosomes at the molecular level. The fourth—by Creagan and Ruddle—is on human gene mapping using new methods, particularly by means of man/rodent somatic cell hybrids. This is by far the best chapter of the book for comprehensiveness, clarity, quality of tabulated information, and soundness of deductions. But what has it to do with the title of the book? A further three chapters, by Bahr, Rao and Dutrillaux, respectively, cover mainly the structure of chromosomes at levels between electron and light microscopy. Finally, chapters 8 and 9, by Pearson and Jones, respectively, deal with the comparative chromosomology of primates—the former on the basis of banding patterns and gene homologies, the latter on the basis of the kinds and distributions of repetitive DNAs.

Processing of heterogenous nuclear RNA, the bulk of which never gets out of the nucleus; chromosomal localisation of specific DNA sequences; human satellite DNAs; human gene localisation using RNA: DNA *in situ* annealing (I dislike the term “hybridisation”) are all treated extensively and competently in the chapters on the molecular level of chromosome organisation. For readers who are not in this field, the treatment illuminates what has been done and how, and what could be done. It is still too soon to draw even a preliminary model on how chromosome structure relates to function.

Chapter 5, by Bahr on “Chromosomes and Chromatin structure”, deals mainly with the 200-Å chromatin fibres, their folding and unfolding, and their relationship to chromomeres, to sister chromatid exchanges and to re-arrangements. The low magnification electron

micrographs are fascinating. To the interpretations one could apply, for the time being, the Italian saying “se son rose fioriranno” (“if there are roses they will bloom”).

Chapter 6, by Rao on premature chromosome condensation, (PCC), shows how an incidental observation made in the early stages of the study of somatic cell fusion led to interesting developments. PCC occurs in the chromosomes of an interphase cell when it fuses with a mitotic cell. It opens the way to the study, in interphase chromosomes, of the distribution of replicating segments and of some aspects of the induction of chromosome aberrations by radiation.

The introduction by Caspersson and his colleagues in 1970 of the banding techniques into human cytology has fostered a flood of major and minor variations: some add to the resolution of the human chromosomes, and others to the confusion. Chapter 7, by Dutrillaux on “New Chromosome Techniques” contributes more to the latter than to the former, in part at least

on account of poor presentation. It does give, however, a good account on the quenching effect of BUdR incorporation on stainability, and its consequences on producing banding patterns, on making sister chromatid exchanges detectable, and so on.

The final two chapters are stimulating, clear, well written and informative. They discuss the light that banding patterns, distributions of homologous genes and distributions of repetitive DNAs throw on primate evolution.

This book is likely to be useful to outsiders or to people working on one approach to human chromosome structure and function, who wish to know something about other approaches. Abundantly obvious is its lack of unity in style, level and coverage, typical of the present plague of multi-author publications based on editorship without dictatorship.

G. Pontecorvo

G. Pontecorvo was on the research staff of the Imperial Cancer Research Fund, London, until his retirement in 1975.

Liquid crystalline state

Liquid Crystals. By S. Chandrasekhar. Pp. 352. (Cambridge University: Cambridge, London and New York, 1977.) £18.

EVERYONE has now heard of liquid crystals and seen a liquid crystal display device. Few know very much about these materials, even physical scientists, although this situation is now changing quite rapidly. The liquid crystalline state of matter is very common—the fact that the number of known examples is not more than several thousand is probably only because more have not been sought or many have not been recognised. The past decade has seen a great upsurge of work in this area, particularly because of the increasing use of the materials in display devices but also because the time was ripe from a purely scientific point of view.

The use of the name liquid crystal is sometimes criticised and the word ‘Mesophase’ is often preferred. The hybrid name, however, is evocative of the mystery and attraction of the materials, but the mongrel offspring is more than simply a half of each parent. Indeed, both the liquid and the crystal aspects of behaviour are strong, but much of the interest and the usefulness of the materials resides in the essentially hybrid properties.

As is usual in a field of research at this stage of development, a variety of textbooks and monographs have recently appeared and there is necessarily some

overlap among them. Professor Chandrasekhar's monograph is a valuable addition to this still small collection. It inevitably invites comparison on the one hand with the introductory textbook by Priestley, Wojtowicz and Sheng, and on the other hand with the elegant and unique monograph on the *Physics of Liquid Crystals* by de Gennes. They all have their place, in that the coverage, approach and level of treatment are significantly different.

Chandrasekhar's book is essentially an extended critical review, primarily of nematics and cholesterics. After a brief introduction (10 pages) there follows a chapter on ‘Statistical Theories of Nematic Order’ (84 pages) and ‘Continuum Theory of the Nematic State’ (89 pages). There is an extensive discussion of Cholesteric Liquid Crystals (88 pages) and the final chapter on Smectic Liquid Crystals is the shortest (47 pages). This deals only with the smectic A and smectic C phases and does not reflect the extent of current interest and activity in the various smectic phases, but perhaps does reflect the current state of detailed understanding of them.

This is a physicist's book rather than a chemist's, and is certainly not one for beginners. One of the difficulties of working in this area is that many disciplines must be brought to bear on a problem and few scientists can be expert enough in all. In this respect Chandrasekhar's book will be a valuable addition to the libraries of those already actively working on liquid crystals as well as those contemplating the adventure.

A. J. Leadbetter

A. J. Leadbetter is Professor of Physical Chemistry at the University of Exeter, UK.

Basic processes in protein biosynthesis

Molecular Mechanisms of Protein Biosynthesis. Edited by Herbert Weissbach and Sidney Pestka. Pp. 720. (Academic: New York and London, 1977.) £39.05; \$55.

In the mid-1960s, the genetic code had been solved and most of the basic processes in protein biosynthesis defined. During the past ten years, knowledge of the detailed mechanisms has grown enormously by contributions from different disciplines and this is reviewed in the collection of 13 articles edited by Weissbach and Pestka.

As Lipmann remarks in his introduction, the pairing of nucleic acid bases represents the driving force in genetic information transfer. The actual mechanism of protein biosynthesis, however, depends to a larger extent on interactions between nucleic acids and proteins. Of the components involved, the structure of only one, transfer RNA, is known. Consequently, ingenious indirect methods have been used to study molecular interactions in protein biosynthesis and many of the chapters are concerned with these.

The complexity of the ribosome, which in *Escherichia coli* is assembled from 54 different proteins and three RNA molecules, poses formidable problems. These are clearly defined in a stimulating article by Kurland, which serves as an essential introduction for many of the other chapters. The use of cross-linking reagents in determining the relative locations of ribosomal components is also discussed here.

Stöffler and Wittmann describe the ribosomal proteins and studies of their location on the ribosomal surface by electron microscopic visualisation of the attachment of antibodies to individual proteins. This very ingenious method is at present somewhat limited by the resolution of the ribosome in the microscope. The statement that "figuratively, the structure of the 30S subunit resembles an embryo or a telephone receiver" is an attempt to convey an impression of the shape, but it underscores certain ambiguities in resolution. Knowledge of ribosome structure may soon improve, probably from studies of crystalline arrays.

More than 200 genes are involved in the specification of the protein synthetic apparatus, and genetic analysis has played a key role in identifying these components. I. Smith gives an excellent discussion of the genetics and control of biosynthesis of the translational system. Pestka contributes a comprehensive review of protein synthesis inhibitors, principally antibiotics. The use of the latter, in conjunction with

genetics and biochemistry, has been the most powerful method in discriminating individual steps in protein biosynthesis. Even in such a complicated system, affinity labelling has been successful in exploring the functional topology of components, and the critical review by Pellegrini and Cantor will interest anyone using these techniques.

Other chapters discuss transfer RNA and synthetases, messenger RNA and individual steps in protein synthesis. Unfortunately, the volume does not include an account of the *in vitro* assembly of the ribosome, which has played an important role in understanding its structural function.

With all this information, it is remarkable that many of the basic mechanisms in protein synthesis still elude us—notably, the fundamental translocation process by which the mRNA moves relative to the ribosome. The answers to this and other questions

depends on a still more difficult task—study of conformational changes in the ribosome.

In their choice of topics, the editors have skilfully encompassed the field. The chapters overlap to some extent, and cross-referencing would have been useful. Because they are simply designated by numbers, the ribosomal proteins are particularly difficult to follow in discussion, and a tabulation of their properties as an appendix would have greatly helped. In most of the chapters the discussion is set in a good conceptual framework although a few are simply reviews.

The book is a unique presentation of the present situation in this complex field and fulfills an essential need for those involved.

J. D. Smith

J. D. Smith is on the staff of the MRC Laboratory of Molecular Biology, Cambridge, UK.

Classic work of modern physics

Quantum Mechanics of One- and Two-Electron Atoms. By Hans A. Bethe and Edwin E. Salpeter. Pp. 369. (Plenum/Rosetta: New York, 1977.) \$8.95.

This work by Bethe and Salpeter first appeared in 1956 in the *Encyclopaedia of Physics*, edited by S. Flügge, and was re-issued in 1957 as a separate book published by Springer. The present paperback edition is changed only by the correction of a few misprints and by the addition of new references where they correct actual errors. The publishers describe it, quite correctly, as "one of the classic works of modern physics", and they are to be congratulated on making it available at a very reasonable price.

The theory of the hydrogen atom, in the Schrödinger and Dirac forms, provides one of the exactly soluble problems of quantum mechanics. Comparison of the theory of radiative corrections with precision measurements (Lamb shift) provides crucial tests of quantum electrodynamics. The theory of the helium atom presents problems which are not exactly soluble; treatment of these problems lays the foundations for the development of the theory of many-electron atoms. The work of Bethe and Salpeter is particularly valuable in the discussion of relativistic effects, which are treated in an unsatisfactory way in many textbooks on atomic structures. The book also contains important sections on atoms in external electric and magnetic fields, and on interactions with radiation.

In their preface to the present edition, the authors say that they have

left the book unchanged not because so little has happened, but because so much has happened. This will, I am sure, be welcomed: it is best that there should be no tampering with classics, even by the authors themselves! The book as it stands is of enduring value.

Yet the reviewer is tempted to ask: what has happened in the past 20 years? I would say, first, the development of the theory of continuum states of two-electron systems. Bethe and Salpeter do not exclude discussion of continuum states, in that they give a full discussion of the continuum states of hydrogenic atoms and some discussion of photo-ionisation of two electron atoms. There have, however, been major advances in the theory of electron collisions with one-electron atoms, including studies of elastic and inelastic scattering; doubly-excited states which auto-ionise; and very fundamental problems in the theory of electron-impact ionisation, giving two electrons in the continuum. There are also the problems of radiation by one-electron atoms in external fields due to an added electron, that is to say the theory of spectral line broadening by a perturber electron.

Secondly, there are multi-photon problems. Bethe and Salpeter discuss multi-photon processes such as the $H\ 2s \rightarrow 1s$ decay with emission of two photons but not problems, such as the ionisation of atoms by intense coherent radiation fields, which have become of interest due to developments in laser technology.

Who, I wonder, will write reviews of the later developments with a mastery approaching that of the book by Bethe and Salpeter?

M. J. Seaton

M. J. Seaton is Professor of Physics at University College, London, UK.